

Supplementary Material for Geometry and Symmetry in Short-and-Sparse Deconvolution

Anonymous authors

A Notations

All vectors/matrices are written in bold font \mathbf{a}/\mathbf{A} ; indexed values are written as $\mathbf{a}_i, \mathbf{A}_{ij}$. Zeros or ones vectors are defined as $\mathbf{0}$ or $\mathbf{1}$, and i -th canonical basis vector defined as \mathbf{e}_i . The indices for vectors/matrices all start from 0 and is taking modulo- n , thus a vector of length n should has its indices labeled as $\{0, 1, \dots, n-1\}$. We write $[n] = \{0, \dots, n-1\}$. We often use captial italic symbols I, J for subsets of $[n]$. We abuse notation slightly and write $[-p] = \{n-p+1, \dots, n-1, 0\}$ and $[\pm p] = \{n-p+1, \dots, n-1, 0, 1, \dots, p-1\}$. Index sets can be labels for vectors; $\mathbf{a}_I \in \mathbb{R}^{|I|}$ denotes the restriction of the vector \mathbf{a} to coordinates I . Also, we use check symbol for reversal operator on index set $\check{I} = -I$ and vectors $\check{\mathbf{a}}_i = \mathbf{a}_{-i}$.

We let P_C denote the projection operator associated with a compact set C . The zero-filling operator $\iota_I : \mathbb{R}^{|I|} \rightarrow \mathbb{R}^n$ injects the input vector to higher dimensional Euclidean space, via $(\iota_I \mathbf{x})_i = \mathbf{x}_{I^{-1}(i)}$ for $i \in I$ and 0 otherwise. Its adjoint operator ι_I^* can be understood as subset selection operator which picks up entries of coordinates I . A common zero-filling operator through out this paper ι is abbreviation of $\iota_{[p]}$, which is often being addressed as zero-padding operator and its adjoint ι^* as truncation operator.

The convolution operator are all circular with modulo- n : $(\mathbf{a} * \mathbf{x})_i = \sum_{j \in [n]} \mathbf{a}_j \mathbf{x}_{i-j}$, also, the convolution operator works on index set: $I * J = \text{supp}(\mathbf{1}_I * \mathbf{1}_J)$. Similarly, the shift operator $s_\ell[\cdot] : \mathbb{R}^p \rightarrow \mathbb{R}^n$ is circular with modulo- n without specification: $(s_\ell[\mathbf{a}])_j = (\iota_{[p]} \mathbf{a})_{j-\ell}$. Notice that here \mathbf{a} can be shorter $p \leq n$. Let $\mathbf{C}_\mathbf{a} \in \mathbb{R}^{n \times n}$ denote a circulant matrix (with modulo- n) for vector \mathbf{a} , whose j -th column is the cyclic shift of \mathbf{a} by j : $\mathbf{C}_\mathbf{a} \mathbf{e}_j = s_j[\mathbf{a}]$. It satisfies for any $\mathbf{b} \in \mathbb{R}^n$,

$$\mathbf{C}_\mathbf{a} \mathbf{b} = \mathbf{a} * \mathbf{b}. \quad (\text{A.1})$$

The correlation between \mathbf{a} and \mathbf{b} can be also written in similar form of convolution operator which reverse one vector before convolution. Define two correlation matrices $\mathbf{C}_\mathbf{a}^*$ and $\check{\mathbf{C}}_\mathbf{a}$ as $\mathbf{C}_\mathbf{a}^* \mathbf{e}_j = s_j[\check{\mathbf{a}}]$ and $\check{\mathbf{C}}_\mathbf{a} \mathbf{e}_j = s_{-j}[\mathbf{a}]$. The two operators will satisfy

$$\mathbf{C}_\mathbf{a}^* \mathbf{b} = \check{\mathbf{a}} * \mathbf{b}, \quad \check{\mathbf{C}}_\mathbf{a} \mathbf{b} = \mathbf{a} * \check{\mathbf{b}}. \quad (\text{A.2})$$

B Geometry of φ_ρ in Shift Space

Underlying our main geometric and algorithmic results is a relationship between the geometry of the function φ_ρ and the symmetries of the deconvolution problem. In this section, we present [Theorem B.1](#), which is a more general result of the main geometric theorem [Theorem 3.1](#). We describe this relationship at a more technical level, by interpreting the gradient and hessian of the function φ_ρ in terms of the shifts $s_\ell[\mathbf{a}_0]$ and stating a key lemma which asserts that a certain neighborhood of the union of subspaces $\Sigma_{4\theta\rho_0}$ can be decomposed into regions of negative curvature, strong gradient, and strong convexity near the target solutions $\pm s_\ell[\mathbf{a}_0]$.

B.1 Shifts and Correlations

The set $\Sigma_{4\theta p_0}$ is a union of subspaces. Any point \mathbf{a} in one of these subspaces \mathcal{S}_τ is a superposition of shifts of \mathbf{a}_0 :

$$\mathbf{a} = \sum_{\ell \in \tau} \alpha_\ell s_\ell[\mathbf{a}_0]. \quad (\text{B.1})$$

This representation can be extended to a general point $\mathbf{a} \in \mathbb{S}^{p-1}$ by writing

$$\mathbf{a} = \sum_{\ell \in \tau} \alpha_\ell s_\ell[\mathbf{a}_0] + \sum_{\ell \notin \tau} \alpha_\ell s_\ell[\mathbf{a}_0]. \quad (\text{B.2})$$

The vector α can be viewed as the coefficients of a decomposition of \mathbf{a} into different shifts of \mathbf{a}_0 . This representation is not unique. For \mathbf{a} close to \mathcal{S}_τ , we can choose a particular α for which α_{τ^c} is small, a notion that we will formalize below.

For convenience, we introduce a closely related vector $\beta \in \mathbb{R}^n$, whose entries are the inner products between \mathbf{a} and the shifts of \mathbf{a}_0 : $\beta_\ell = \langle \mathbf{a}, s_\ell[\mathbf{a}_0] \rangle$. Since the columns of $\mathbf{C}_{\mathbf{a}_0}$ are the shifts of \mathbf{a}_0 , we can write

$$\begin{aligned} \beta &= \mathbf{C}_{\mathbf{a}_0}^* \iota \mathbf{a} & (\text{B.3}) \\ &= \mathbf{C}_{\mathbf{a}_0}^* \iota^* \mathbf{C}_{\mathbf{a}_0} \alpha =: \mathbf{M} \alpha. & (\text{B.4}) \end{aligned}$$

The matrix \mathbf{M} is the Gram matrix of the truncated shifts: $M_{ij} = \langle \iota^* s_i[\mathbf{a}_0], \iota^* s_j[\mathbf{a}_0] \rangle$. When μ is small, the off-diagonal elements of \mathbf{M} are small. In particular, on \mathcal{S}_τ we may take $\alpha_{\tau^c} = \mathbf{0}$, and $\beta \approx \alpha$, in the sense that $\beta_\tau \approx \alpha_\tau$ and the entries of β_{τ^c} are small. For detailed elaboration, see [Appendix E](#).

B.2 Shifts and the Calculus of φ_{ℓ^1}

Our main geometric claims pertain to the function φ_ρ , which is based on a smooth sparsity surrogate $\rho(\cdot) \approx \|\cdot\|_1$. In this section, we sketch the main ideas of the proof as if $\rho(\cdot) = \|\cdot\|_1$, by relating the geometry of the function φ_{ℓ^1} to the vectors α, β introduced above. Working with φ_{ℓ^1} simplifies the exposition; it is also faithful to the structure of our proof, which relates the derivatives of the smooth function φ_ρ to similar quantities associated with the nonsmooth function φ_{ℓ^1} .

The function φ_{ℓ^1} has a relatively simple closed form:

$$\varphi_{\ell^1}(\mathbf{a}) = -\frac{1}{2} \|\mathcal{S}_\lambda[\check{\mathbf{y}} * \mathbf{a}]\|_2^2. \quad (\text{B.5})$$

Here, \mathcal{S}_λ is the *soft thresholding operator*, which is defined for scalars t as $\mathcal{S}_\lambda[t] = \text{sign}(t) \max\{|t| - \lambda, 0\}$, and is extended to vectors by applying it elementwise. The operator $\mathcal{S}_\lambda[\mathbf{x}]$ shrinks the elements of \mathbf{x} towards zero. Small elements become identically zero, resulting in a sparse vector.

Gradient: Sparsifying the Correlations β

Our goal is to understand the local minimizers of the function φ_{ℓ^1} over the sphere. The function φ_{ℓ^1} is differentiable. Clearly, any point \mathbf{a} at which its gradient (over the sphere) is nonzero cannot be a local minimizer. We first give an expression for the gradient of φ_{ℓ^1} over Euclidean space \mathbb{R}^p , and then extend it to the sphere \mathbb{S}^{p-1} . Using $\mathbf{y} = \mathbf{a}_0 * \mathbf{x}_0$ and calculus gives

$$\begin{aligned} \nabla \varphi_{\ell^1}(\mathbf{a}) &= -\iota^* \mathbf{C}_{\mathbf{a}_0} \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda \left[\check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{C}_{\mathbf{a}_0}^* \iota \mathbf{a} \right] \\ &= -\iota^* \mathbf{C}_{\mathbf{a}_0} \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda \left[\check{\mathbf{C}}_{\mathbf{x}_0} \beta \right] \\ &= -\iota^* \mathbf{C}_{\mathbf{a}_0} \chi[\beta], \end{aligned} \quad (\text{B.6})$$

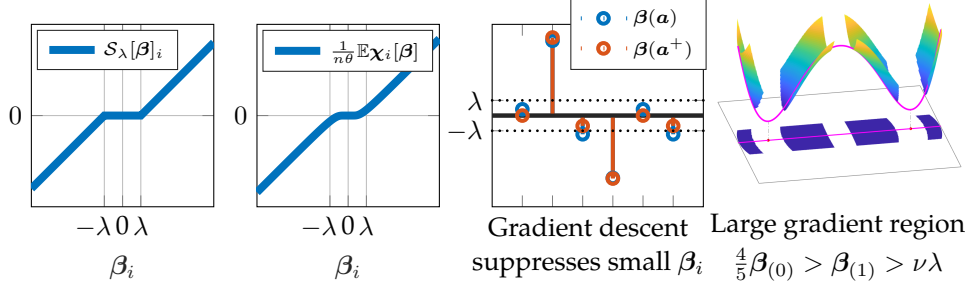


Figure 1: Gradient Sparsifies Correlations. Left: the soft thresholding operator $S_\lambda[\beta]$ shrinks the entries of β towards zero, making it sparser. Middle left: the negative gradient $-\nabla\varphi_{\ell 1}$ is a superposition of shifts $s_\ell[\mathbf{a}_0]$, with coefficients $\chi_\ell[\beta] \approx S_\lambda[\beta]_\ell$. Because of this, gradient descent sparsifies β . Middle right: $\beta(\mathbf{a})$ before, and $\beta(\mathbf{a}^+)$ after, one projected gradient step $\mathbf{a}^+ = \mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{a} - t \cdot \text{grad}[\varphi_{\ell 1}](\mathbf{a})]$. Notice that the small entries of β are shrunk towards zero. Right: the gradient $\text{grad}[\varphi_{\ell 1}](\mathbf{a})$ is large whenever it is easy to sparsify β ; in particular, when the largest entry $\beta_{(0)} \gg \beta_{(1)} \gg 0$.

where we have simplified the notation by introducing an operator $\chi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ as $\chi[\beta] = \check{\mathbf{C}}_{\mathbf{x}_0} S_\lambda[\check{\mathbf{C}}_{\mathbf{x}_0} \beta]$. This representation exhibits the (negative) gradient as a superposition of shifts of \mathbf{a}_0 with coefficients given by the entries of $\chi[\beta]$:

$$-\nabla\varphi_{\ell 1}(\mathbf{a}) = \sum_{\ell} \chi[\beta]_{\ell} s_{\ell}[\mathbf{a}_0]. \quad (\text{B.7})$$

The operator χ appears complicated. However, its effect is relatively simple: *when \mathbf{x}_0 is a long random vector, $\chi[\beta]$ acts like a soft thresholding operator on the vector β .* That is,

$$\frac{1}{n\theta} \cdot \chi[\beta]_{\ell} \approx \begin{cases} \beta_{\ell} - \lambda, & \beta_{\ell} > \lambda \\ \beta_{\ell} + \lambda, & \beta_{\ell} < -\lambda \\ 0, & \text{otherwise} \end{cases}. \quad (\text{B.8})$$

We show this rigorously below, in the proof of our main theorems. Here, we support this claim pictorially, by plotting the ℓ -th entry $\chi[\beta]_{\ell}$ as β_{ℓ} varies – see Figure 1 (middle left) and compare to Figure 1 (left). Because $\chi[\beta]$ suppresses small entries of β , the strongest contributions to $-\nabla\varphi_{\ell 1}$ in (B.7) will come from shifts $s_{\ell}[\mathbf{a}_0]$ with large β_{ℓ} . *In particular, the Euclidean gradient is large whenever there is a single preferred shift $s_{\ell}[\mathbf{a}_0]$, i.e., the largest entry of β is significantly larger than the second largest entry.*

The (Euclidean) gradient $\nabla\varphi_{\ell 1}$ measures the slope of $\varphi_{\ell 1}$ over \mathbb{R}^n . We are interested in the slope of $\varphi_{\ell 1}$ over the sphere \mathbb{S}^{p-1} , which is measured by the Riemannian gradient

$$\begin{aligned} \text{grad}[\varphi_{\ell 1}](\mathbf{a}) &= \mathbf{P}_{\mathbf{a}^\perp} \nabla\varphi_{\ell 1}(\mathbf{a}) \\ &= -\mathbf{P}_{\mathbf{a}^\perp} \sum_{\ell} \chi[\beta]_{\ell} s_{\ell}[\mathbf{a}_0]. \end{aligned} \quad (\text{B.9})$$

The Riemannian gradient simply projects the Euclidean gradient onto the tangent space \mathbf{a}^\perp to \mathbb{S}^{p-1} at \mathbf{a} . The Riemannian gradient is large whenever

- (i) **Negative gradient points to one particular shift:** there is a single preferred shift $s_{\ell}[\mathbf{a}_0]$ so that the Euclidean gradient is large *and*
- (ii) **\mathbf{a} is not too close to any shift:** it is possible to move in the tangent space in the direction of this shift.¹ Since the tangent space consists of those vectors orthogonal to \mathbf{a} , this is possible whenever $s_{\ell}[\mathbf{a}_0]$ is not too aligned with \mathbf{a} , i.e., \mathbf{a} is not too close to $s_{\ell}[\mathbf{a}_0]$.

¹...so the projection of the Euclidean gradient onto the tangent space does not vanish.

Our technical lemma quantifies this situation in terms of the ordered entries of β . Write $|\beta_{(0)}| \geq |\beta_{(1)}| \geq \dots$, with corresponding shifts $s_{(0)}[\mathbf{a}_0], s_{(1)}[\mathbf{a}_0], \dots$. There is a strong gradient whenever $|\beta_{(0)}|$ is significantly larger than $|\beta_{(1)}|$ and $|\beta_{(1)}|$ is not too small compared to λ : in particular, when $\frac{4}{5}|\beta_{(0)}| > |\beta_{(1)}| > \frac{\lambda}{4 \log^2 \theta^{-1}}$. In this situation, gradient descent drives \mathbf{a} toward $s_{(0)}[\mathbf{a}_0]$, reducing $|\beta_{(1)}|, \dots$, and making the vector β sparser. We establish the technical claim that the (Euclidean) gradient of φ_{ℓ^1} sparsifies vectors in shift space in [Appendix F](#).

Hessian: Negative Curvature Breaks Symmetry

When there is no single preferred shift, i.e., when $|\beta_{(1)}|$ is close to $|\beta_{(0)}|$, the gradient can be small. Similarly, when \mathbf{a} is very close to $\pm s_{(0)}[\mathbf{a}_0]$, the gradient can be small. In either of these situations, we need to study the curvature of the function φ to determine whether there are local minimizers.

Strictly speaking, the function φ_{ℓ^1} is not twice differentiable, due to the nonsmoothness of the soft thresholding operator $\mathcal{S}_\lambda[t]$ at $t = \pm\lambda$. Indeed, φ_{ℓ^1} is nonsmooth at any point \mathbf{a} for which some entry of $\check{\mathbf{y}} * \mathbf{a}$ has magnitude λ . At other points \mathbf{a} , φ_{ℓ^1} is twice differentiable, and its Hessian is given by

$$\tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) = -\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \check{\mathbf{C}}_{x_0} \mathbf{P}_I \check{\mathbf{C}}_{x_0}^* \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}, \quad (\text{B.10})$$

with $I = \text{supp}(\check{\mathcal{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a})$. We (formally) extend this expression to *every* $\mathbf{a} \in \mathbb{R}^n$, terming $\tilde{\nabla}^2 \varphi_{\ell^1}$ the *pseudo-Hessian* of φ_{ℓ^1} . For appropriately chosen smooth sparsity surrogate ρ , we will see that the (true) Hessian of the smooth function $\nabla^2 \varphi_\rho$ is close to $\tilde{\nabla}^2 \varphi_{\ell^1}$, and so $\tilde{\nabla}^2 \varphi_{\ell^1}$ yields useful information about the curvature of φ_ρ .

As with the gradient, the Hessian is complicated, but becomes simpler when the sample size is large. The following approximation

$$\tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \approx - \sum_{\ell} s_{\ell}[\mathbf{a}_0] s_{\ell}[\mathbf{a}_0]^* \left(\frac{\partial}{\partial \beta_{\ell}} \boldsymbol{\chi}_{\ell}[\boldsymbol{\beta}] \right) \quad (\text{B.11})$$

can be obtained from [\(B.7\)](#) noting that $\frac{\partial}{\partial \mathbf{a}} \boldsymbol{\chi}_{\ell}[\boldsymbol{\beta}] = \sum_j s_j[\mathbf{a}_0] \frac{\partial}{\partial \beta_j} \boldsymbol{\chi}_{\ell}[\boldsymbol{\beta}]$, that $\frac{\partial}{\partial \beta_j} \boldsymbol{\chi}_{\ell}[\boldsymbol{\beta}] \approx 0$ for $j \neq \ell$, and that

$$\frac{1}{n\theta} \cdot \frac{\partial \boldsymbol{\chi}_{\ell}[\boldsymbol{\beta}]}{\partial \beta_{\ell}} \approx \begin{cases} 0 & |\beta_{\ell}| \ll \lambda \\ 1 & |\beta_{\ell}| \gg \lambda \end{cases} \quad (\text{B.12})$$

Again, we corroborate this approximation pictorially – see [Figure 2](#).

From this approximation, we can see that the quadratic form $\mathbf{v}^* \tilde{\nabla}^2 \varphi_{\ell^1} \mathbf{v}$ takes on a large negative value whenever \mathbf{v} is a shift $s_{\ell}[\mathbf{a}_0]$ corresponding to some $|\beta_{\ell}| \geq \lambda$, or whenever \mathbf{v} is a linear combination of such shifts. *In particular, if for some j , $|\beta_{(0)}|, |\beta_{(1)}|, \dots, |\beta_{(j)}| \gg \lambda$, then φ_{ℓ^1} will exhibit negative curvature in any direction $\mathbf{v} \in \text{span}(s_{(0)}[\mathbf{a}_0], s_{(1)}[\mathbf{a}_0], \dots, s_{(j)}[\mathbf{a}_0])$.*

The (Euclidean) Hessian measures the curvature of the function φ_{ℓ^1} over \mathbb{R}^n . The Riemannian Hessian

$$\widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) = \mathbf{P}_{\mathbf{a}^\perp} \left(\underbrace{\tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a})}_{\text{Curvature of } \varphi_{\ell^1}} + \underbrace{\langle -\nabla \varphi_{\ell^1}(\mathbf{a}), \mathbf{a} \rangle \cdot \mathbf{I}}_{\text{Curvature of the sphere}} \right) \mathbf{P}_{\mathbf{a}^\perp}. \quad (\text{B.13})$$

measures the curvature of φ_{ℓ^1} over the sphere. The projection $\mathbf{P}_{\mathbf{a}^\perp}$ restricts its action to directions $\mathbf{v} \perp \mathbf{a}$ that are tangent to the sphere. The additional term $\langle -\nabla \varphi_{\ell^1}(\mathbf{a}), \mathbf{a} \rangle$ accounts for the curvature of the sphere. This term is always positive. The net effect is that directions of strong negative curvature of φ_{ℓ^1} over \mathbb{R}^n become directions of moderate negative curvature over the sphere. Directions of nearly zero curvature over \mathbb{R}^n become directions of positive curvature over the sphere. This has three implications for the geometry of φ_{ℓ^1} over the sphere:

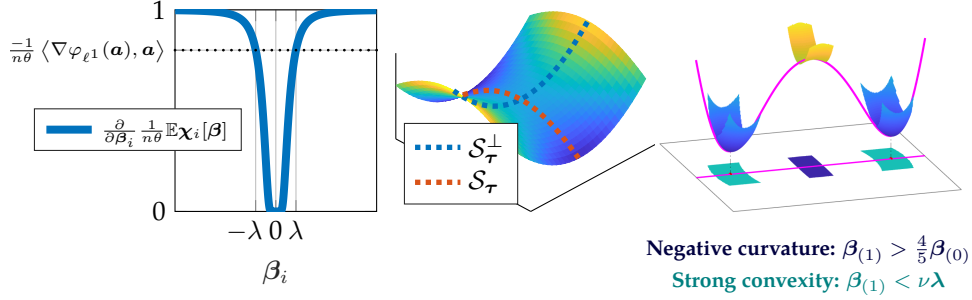


Figure 2: Hessian Breaks Symmetry. Left: contribution of $-s_i[\mathbf{a}_0]s_i[\mathbf{a}_0]^*$ to the Euclidean Hessian. If $|\beta_i| \gg \lambda$ the Euclidean Hessian exhibits a strong negative component in the $s_i[\mathbf{a}_0]$ direction. The Riemannian Hessian exhibits negative curvature in directions spanned by $s_i[\mathbf{a}_0]$ with corresponding $|\beta_i| \gg \lambda$ and positive curvature in directions spanned by $s_i[\mathbf{a}_0]$ with $|\beta_i| \ll \lambda$. Middle: this creates negative curvature along the subspace \mathcal{S}_τ and positive curvature orthogonal to this subspace. Right: our analysis shows that there is always a direction of negative curvature when $\beta_{(1)} > \frac{4}{5}\beta_{(0)}$; conversely when $\beta_{(1)} \ll \lambda$ there is positive curvature in every feasible direction and the function is strongly convex.

- (i) **Negative curvature in symmetry breaking directions:** If $|\beta_{(0)}|, |\beta_{(1)}|, \dots, |\beta_{(j)}| \gg \lambda$, φ_{ℓ^1} will exhibit negative curvature in any tangent direction $\mathbf{v} \perp \mathbf{a}$ which is in the linear span

$$\text{span}(s_{(0)}[\mathbf{a}_0], s_{(1)}[\mathbf{a}_0], \dots, s_{(j)}[\mathbf{a}_0])$$

of the corresponding shifts of \mathbf{a}_0 .

- (ii) **Positive curvature in directions away from \mathcal{S}_τ :** The Euclidean Hessian quadratic form $\mathbf{v}^* \tilde{\nabla}^2 \varphi_{\ell^1} \mathbf{v}$ takes on relatively small values in directions orthogonal to the subspace \mathcal{S}_τ . The Riemannian Hessian is positive in these directions, creating positive curvature orthogonal to the subspace \mathcal{S}_τ .
- (iii) **Strong convexity around minimizers:** Around a minimizer $s_\ell[\mathbf{a}_0]$, only a single entry β_ℓ is large. Any tangent direction $\mathbf{v} \perp \mathbf{a}$ is nearly orthogonal to the subspace $\text{span}(s_\ell[\mathbf{a}_0])$, and hence is a direction of positive (Riemannian) curvature. The objective function φ_ρ is strongly convex around the target solutions $\pm s_\ell[\mathbf{a}_0]$.

Figure 2 visualizes these regions of negative and positive curvature, and the technical claim of positivity/negativity of curvature in shift space is presented in detail in Appendix G.

B.3 Any Local Minimizer is a Near Shift

We close this section by stating a key theorem, which makes the above discussion precise. We will show that a certain neighborhood of any subspace \mathcal{S}_τ can be covered by regions of *negative curvature, large gradient*, and regions of *strong convexity* containing target solutions $\pm s_\ell[\mathbf{a}_0]$. Furthermore, at the boundary of this neighborhood, the negative gradient points back—*retracts*—toward the subspace \mathcal{S}_τ , due to the (directional) convexity of φ_ρ away from the subspace.

To formally state the result, we need a way of measuring how close \mathbf{a} is to the subspace \mathcal{S}_τ . For technical reasons, it turns out to be convenient to do this in terms of the coefficients α in the representation

$$\mathbf{a} = \sum_{\ell \in \tau} \alpha_\ell s_\ell[\mathbf{a}_0] + \sum_{\ell' \in \tau^c} \alpha_{\ell'} s_{\ell'}[\mathbf{a}_0]. \quad (\text{B.14})$$

If $\mathbf{a} \in \mathcal{S}_\tau$, we can take $\alpha_{\tau^c} = 0$. We can view the energy $\|\alpha_{\tau^c}\|_2$ as a measure of the distance from \mathbf{a} to \mathcal{S}_τ . A technical wrinkle arises, because the representation (B.14) is not unique. We resolve this issue by

choosing the α that minimizes $\|\alpha_{\tau^c}\|_2$, writing:

$$d_\alpha(\mathbf{a}, \mathcal{S}_\tau) = \inf \{ \|\alpha_{\tau^c}\|_2 : \sum_\ell \alpha_\ell s_\ell[\mathbf{a}_0] = \mathbf{a} \}. \quad (\text{B.15})$$

The distance $d_\alpha(\mathbf{a}, \mathcal{S}_\tau)$ is zero for $\mathbf{a} \in \mathcal{S}_\tau$. Our analysis controls the geometric properties of φ_ρ over the set of \mathbf{a} for which $d_\alpha(\mathbf{a}, \mathcal{S}_\tau)$ is not too large. Similar to (??), we define an object which contains all points that are close to some \mathcal{S}_τ , in the above sense:

$$\Sigma_{4\theta p_0}^\gamma := \bigcup_{|\tau| \leq 4\theta p_0} \{ \mathbf{a} : d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \leq \gamma \}. \quad (\text{B.16})$$

The aforementioned geometric properties hold over this set:

Theorem B.1 (Three subregions). *Suppose that $\mathbf{y} = \mathbf{a}_0 * \mathbf{x}_0$ where $\mathbf{a}_0 \in \mathbb{S}^{p_0-1}$ is μ -shift coherent and $\mathbf{x}_0 \sim \text{i.i.d.}$ $\text{BG}(\theta) \in \mathbb{R}^n$ satisfying*

$$\theta \in \left[\frac{c'}{p_0}, \frac{c}{p_0 \sqrt{\mu} + \sqrt{p_0}} \right] \cdot \frac{1}{\log^2 p_0} \quad (\text{B.17})$$

for some constants $c', c > 0$. Set $\lambda = 0.1/\sqrt{p_0\theta}$ in φ_ρ where $\rho(x) = \sqrt{x^2 + \delta^2}$. There exist numerical constants $C, c'', c''', c_1-c_4 > 0$ such that if $\delta \leq \frac{c''\lambda\theta^8}{p^2 \log^2 n}$ and $n > Cp_0^5 \theta^{-2} \log p_0$, then with probability at least $1 - c'''/n$, for every $\mathbf{a} \in \Sigma_{4\theta p_0}^\gamma$, we have:

- (Negative curvature): If $|\beta_{(1)}| \geq \nu_1 |\beta_{(0)}|$, then

$$\lambda_{\min}(\text{Hess}[\varphi_\rho](\mathbf{a})) \leq -c_1 n \theta \lambda; \quad (\text{B.18})$$

- (Large gradient): If $\nu_1 |\beta_{(0)}| \geq |\beta_{(1)}| \geq \nu_2(\theta)\lambda$, then

$$\|\text{grad}[\varphi_\rho](\mathbf{a})\|_2 \geq c_2 n \theta \frac{\lambda^2}{\log^2 \theta^{-1}}; \quad (\text{B.19})$$

- (Convex near shifts): If $\nu_2(\theta)\lambda \geq |\beta_{(1)}|$, then

$$\text{Hess}[\varphi_\rho](\mathbf{a}) \succ c_3 n \theta \mathbf{P}_{\mathbf{a}^\perp}; \quad (\text{B.20})$$

- (Retraction to subspace): If $\frac{\gamma}{2} \leq d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \leq \gamma$, then for every α satisfying $\mathbf{a} = \iota^* \mathbf{C}_{\mathbf{a}_0} \alpha$, there exists ζ satisfying $\text{grad}[\varphi_\rho](\mathbf{a}) = \iota^* \mathbf{C}_{\mathbf{a}_0} \zeta$, such that

$$\langle \zeta_{\tau^c}, \alpha_{\tau^c} \rangle \geq c_4 \|\zeta_{\tau^c}\|_2 \|\alpha_{\tau^c}\|_2; \quad (\text{B.21})$$

- (Local minimizers): If \mathbf{a} is a local minimizer,

$$\min_{\substack{\ell \in \{\pm p\} \\ \sigma \in \{\pm 1\}}} \|\mathbf{a} - \sigma s_\ell[\mathbf{a}_0]\|_2 \leq \frac{1}{2} \max \{ \mu, p_0^{-1} \}, \quad (\text{B.22})$$

where $\nu_1 = \frac{4}{5}$, $\nu_2(\theta) = \frac{1}{4 \log^2 \theta^{-1}}$ and $\gamma = \frac{c \cdot \text{poly}(\sqrt{1/\theta}, \sqrt{1/\mu})}{\log^2 \theta^{-1}} \cdot \frac{1}{\sqrt{p_0}}$.

Proof. See [Appendix I.5](#). ■

The retraction property elaborated in (B.21) implies that the negative gradient at \mathbf{a} points in a direction that decreases $d_\alpha(\mathbf{a}, \mathcal{S}_\tau)$. This is a consequence of positive curvature away from \mathcal{S}_τ . It essentially implies that the gradient is monotone in α_{τ^c} space: choose any $\underline{\mathbf{a}} \in \mathcal{S}_\tau \cap \mathbb{S}^{p-1}$, write $\underline{\alpha}$ to be its coefficient, and let $\underline{\zeta}$ be the coefficient of $\text{grad}[\varphi_\rho](\underline{\mathbf{a}})$. Then $\underline{\alpha}_{\tau^c} = \mathbf{0}$, $\underline{\zeta}_{\tau^c} \approx \mathbf{0}$ and

$$\langle \zeta_{\tau^c} - \underline{\zeta}_{\tau^c}, \alpha_{\tau^c} - \underline{\alpha}_{\tau^c} \rangle \approx \langle \zeta_{\tau^c} - \mathbf{0}, \alpha_{\tau^c} - \mathbf{0} \rangle = \langle \zeta_{\tau^c}, \alpha_{\tau^c} \rangle > 0.$$

Our main geometric claim [Theorem 3.1](#) is a direct consequence of [Theorem B.1](#). Moreover, it suggests that as long as we can minimize φ_ρ within the region $\Sigma_{4\theta p_0}^\gamma$, we will solve the SaS deconvolution problem.

C Provable Algorithm

In light of [Theorem B.1](#), we introduced a two-part algorithm [Algorithm 1](#), which first applies the curvilinear descent method to find a local minimum of φ_ρ within $\Sigma_{4\theta p_0}^\gamma$, followed by refinement algorithm that uses alternating minimization to exactly recover the ground truth. This algorithm exactly solves SaS deconvolution problem. In this section, we demonstrate [Theorem C.1](#) and [Theorem C.2](#), which are sufficient to prove [Theorem 3.2](#) when jointly combined.

C.1 Minimization

There are three major issues in finding a local minimizer within $\Sigma_{4\theta p_0}^\gamma$. We want ...

- (i) **Initialization.** the initializer $\mathbf{a}^{(0)}$ to reside within $\Sigma_{4\theta p_0}^\gamma$,
- (ii) **Negative curvature.** the method to avoid stagnating near the saddle points of φ_ρ ,
- (iii) **No exit.** the descent method to remain inside $\Sigma_{4\theta p_0}^\gamma$.

In the following paragraphs, we describe how our proposed algorithm achieves the above desiderata.

Initialization within $\Sigma_{4\theta p_0}^\gamma$. Our data-driven initialization scheme produces $\mathbf{a}^{(0)}$, where

$$\begin{aligned}\mathbf{a}^{(0)} &= -\mathbf{P}_{\mathbb{S}^{p-1}} \nabla \varphi_\rho (\mathbf{P}_{\mathbb{S}^{p-1}} [\mathbf{0}^{p_0-1}; \mathbf{y}_0; \dots; \mathbf{y}_{p_0-1}; \mathbf{0}^{p_0-1}]) \\ &= -\mathbf{P}_{\mathbb{S}^{p-1}} \nabla \varphi_\rho \mathbf{P}_{\mathbb{S}^{p-1}} [\mathbf{P}_{[p_0]}(\mathbf{a}_0 * \mathbf{x}_0)], \\ &\approx -\mathbf{P}_{\mathbb{S}^{p-1}} \nabla \varphi_\rho [\mathbf{P}_{[p_0]}(\mathbf{a}_0 * \tilde{\mathbf{x}}_0)],\end{aligned}$$

is the normalized gradient vector from a chunk of data $\mathbf{a}^{(-1)} := \mathbf{P}_{[p_0]}(\mathbf{a}_0 * \tilde{\mathbf{x}}_0)$ with $\tilde{\mathbf{x}}_0$ a normalized Bernoulli-Gaussian random vector of length $2p_0 - 1$. Since $\nabla \varphi_\rho \approx \nabla \varphi_{\ell^1}$, expand the gradient $\nabla \varphi_{\ell^1}$ and rewrite the gradient $\nabla_{\ell^1}(\mathbf{a}^{(-1)})$ in shift space, we get

$$\begin{aligned}-\nabla \varphi_{\rho^1}(\mathbf{a}^{(-1)}) &\approx \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda [\check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_{[p_0]}(\mathbf{a}_0 * \tilde{\mathbf{x}}_0)] \\ &= \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\chi} [\mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_{[p_0]} \mathbf{C}_{\mathbf{a}_0} \tilde{\mathbf{x}}_0] \\ &\approx \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\chi} [\tilde{\mathbf{x}}_0] \\ &\approx n\theta \cdot \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \mathcal{S}_\lambda [\tilde{\mathbf{x}}_0],\end{aligned}$$

where the approximation in the third equation is accurate if the truncated shifts are incoherent

$$\max_{i \neq j} |\langle \boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0], \boldsymbol{\iota}_{p_0}^* s_j[\mathbf{a}_0] \rangle| \leq \mu \ll 1. \quad (\text{C.1})$$

With this simple approximation, it comes clear that the coefficients (in shift space) of initializer $\mathbf{a}^{(0)}$,

$$\mathbf{a}^{(0)} \approx \mathbf{P}_{\mathbb{S}^{p-1}} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \mathcal{S}_\lambda [\tilde{\mathbf{x}}_0], \quad (\text{C.2})$$

approximate $\mathcal{S}_\lambda [\tilde{\mathbf{x}}_0]$, which resides near the subspace \mathcal{S}_τ , in which τ contains the nonzero entries of $\tilde{\mathbf{x}}_0$ on $\{-p_0 + 1, \dots, p_0 - 1\}$. With high probability, the number of non-zero entries is $|\tau| \lesssim 4\theta p_0$, we therefore conclude that our initializer $\mathbf{a}^{(0)}$ satisfies

$$\mathbf{a}^{(0)} \in \Sigma_{4\theta p_0}^\gamma. \quad (\text{C.3})$$

Furthermore, since $\tilde{\mathbf{x}}_0$ is normalized, the largest magnitude for entries of $|\tilde{\mathbf{x}}_0|$ is likely to be around $1/\sqrt{2p_0\theta}$. To ensure that $\mathcal{S}_\lambda [\tilde{\mathbf{x}}_0]$ does not annihilate all nonzero entries of $\tilde{\mathbf{x}}_0$ (otherwise our initializer $\mathbf{a}^{(0)}$ will become $\mathbf{0}$), the ideal λ should be slightly less than the largest magnitude of $|\tilde{\mathbf{x}}_0|$. We suggest setting λ in φ_ρ as

$$\lambda = \frac{c}{\sqrt{p_0\theta}}. \quad (\text{C.4})$$

for some $c \in (0, 1)$.

Many methods have been proposed to optimize functions whose saddle points exhibit strict negative curvature, including the noisy gradient method [GHJY15], trust region methods [AMS09, SQW17] and curvilinear search [WY13]. Any of the above methods can be adapted to minimize φ_ρ . In this paper, we use *curvilinear method with restricted stepsize* to demonstrate how to analyze an optimization problem using the geometric properties of φ_ρ over $\Sigma_{4\theta p_0}^\gamma$ – in particular, negative curvature in symmetry-breaking directions and positive curvature away from \mathcal{S}_τ .

Curvilinear search uses an update strategy that combines the gradient \mathbf{g} and a direction of negative curvature \mathbf{v} , which here we choose as an eigenvector of the hessian \mathbf{H} with smallest eigenvalue, scaled such that $\mathbf{v}^* \mathbf{g} \geq 0$. In particular, we set

$$\mathbf{a}^+ \leftarrow \mathbf{P}_{\mathbb{S}^{p-1}} [\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}] \quad (\text{C.5})$$

For small t ,

$$\varphi(\mathbf{a}^+) \approx \varphi(\mathbf{a}) + \langle \mathbf{g}, \boldsymbol{\xi} \rangle + \frac{1}{2} \boldsymbol{\xi}^* \mathbf{H} \boldsymbol{\xi}. \quad (\text{C.6})$$

Since $\boldsymbol{\xi}$ converges to $\mathbf{0}$ only if \mathbf{a} converges to the local minimizer (otherwise either gradient \mathbf{g} is nonzero or there is a negative curvature direction \mathbf{v}), this iteration produces a local minimizer for φ_ρ , whose saddle points near any \mathcal{S}_τ has negative curvature, we just need to ensure all iterates stays near some such subspace. We prove this by showing:

- When $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \leq \gamma$, curvilinear steps move a small distance away from the subspace:

$$|d_\alpha(\mathbf{a}^+, \mathcal{S}_\tau) - d_\alpha(\mathbf{a}, \mathcal{S}_\tau)| \leq \frac{\gamma}{2}. \quad (\text{C.7})$$

- When $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \in [\frac{\gamma}{2}, \gamma]$, curvilinear steps retract toward subspace:

$$d_\alpha(\mathbf{a}^+, \mathcal{S}_\tau) \leq d_\alpha(\mathbf{a}, \mathcal{S}_\tau). \quad (\text{C.8})$$

Together, we can prove that the iterates $\mathbf{a}^{(k)}$ converge to a minimizer, and

$$\forall k = 1, 2, \dots, \quad \mathbf{a}^{(k)} \in \Sigma_{4\theta p_0}^\gamma. \quad (\text{C.9})$$

We conclude this section with the following theorem:

Theorem C.1 (Convergence of retractive curvilinear search). *Suppose signals $\mathbf{a}_0, \mathbf{x}_0$ satisfy the conditions of Theorem B.1, $\theta > 10^3 c/p_0$ ($c > 1$), and \mathbf{a}_0 is μ -truncated shift coherent $\max_{i \neq j} |\langle \mathbf{u}_{p_0}^* s_i[\mathbf{a}_0], \mathbf{u}_{p_0}^* s_j[\mathbf{a}_0] \rangle| \leq \mu$. Write $\mathbf{g} = \text{grad}[\varphi_\rho](\mathbf{a})$ and $\mathbf{H} = \text{Hess}[\varphi_\rho](\mathbf{a})$. When the smallest eigenvalue of \mathbf{H} is strictly smaller than $-\eta_v$, let \mathbf{v} be the unit eigenvector of smallest eigenvalue, scaled so $\mathbf{v}^* \mathbf{g} \geq 0$; otherwise let $\mathbf{v} = \mathbf{0}$. Define a sequence $\{\mathbf{a}^{(k)}\}_{k \in \mathbb{N}}$ where $\mathbf{a}^{(0)}$ equals (??) and for $k = 1, 2, \dots, K_1$:*

$$\mathbf{a}^{(k+1)} \leftarrow \mathbf{P}_{\mathbb{S}^{p-1}} [\mathbf{a}^{(k)} - t\mathbf{g}^{(k)} - t^2\mathbf{v}^{(k)}] \quad (\text{C.10})$$

with largest $t \in (0, \frac{0.1}{n\theta}]$ satisfying Armijo steplength:

$$\varphi_\rho(\mathbf{a}^{(k+1)}) < \varphi_\rho(\mathbf{a}^{(k)}) - \frac{1}{2} \left(t \|\mathbf{g}^{(k)}\|_2^2 + \frac{1}{2} t^4 \eta_v \|\mathbf{v}^{(k)}\|_2^2 \right), \quad (\text{C.11})$$

then with probability at least $1 - 1/c$, there exists some signed shift $\bar{\mathbf{a}} = \pm s_i[\mathbf{a}_0]$ where $i \in [\pm p_0]$ such that $\|\mathbf{a}^{(k)} - \bar{\mathbf{a}}\|_2 \leq \mu + 1/p$ for all $k \geq K_1 = \text{poly}(n, p)$. Here, $\eta_v = c'n\theta\lambda$ for some $c' < c_1$ in Theorem B.1.

Proof. See Appendix J.2. ■

C.2 Local Refinement

In this section, we describe and analyze an algorithm which refines an estimate $\bar{\mathbf{a}} \approx \mathbf{a}_0$ of the kernel to exactly recover $(\mathbf{a}_0, \mathbf{x}_0)$. Set

$$\mathbf{a}^{(0)} \leftarrow \bar{\mathbf{a}}, \quad \lambda^{(0)} \leftarrow C(p\theta + \log n)(\mu + 1/p), \quad I^{(0)} \leftarrow \text{supp}(\mathcal{S}_\lambda[\mathbf{C}_{\bar{\mathbf{a}}}^* \mathbf{y}]). \quad (\text{C.12})$$

We alternatively minimize the Lasso objective with respect to \mathbf{a} and \mathbf{x} :

$$\mathbf{x}^{(k+1)} \leftarrow \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2} \|\mathbf{a}^{(k)} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda^{(k)} \sum_{i \notin I^{(k)}} |\mathbf{x}_i|, \quad (\text{C.13})$$

$$\mathbf{a}^{(k+1)} \leftarrow \mathbf{P}_{\mathbb{S}^{p-1}} \left[\underset{\mathbf{a}}{\text{argmin}} \frac{1}{2} \|\mathbf{a} * \mathbf{x}^{(k+1)} - \mathbf{y}\|_2^2 \right], \quad (\text{C.14})$$

$$\lambda^{(k+1)} \leftarrow \frac{1}{2} \lambda^{(k)}, \quad I^{(k+1)} \leftarrow \text{supp}(\mathbf{x}^{(k+1)}). \quad (\text{C.15})$$

One departure from standard alternating minimization procedures is our use of a continuation method, which (i) decreases λ and (ii) maintains a running estimate $I^{(k)}$ of the support set. Our analysis will show that $\mathbf{a}^{(k)}$ converges to one of the signed shifts of \mathbf{a}_0 at a linear rate, in the sense that

$$\min_{\sigma \in \pm 1, \ell \in [\pm p_0]} \|\mathbf{a}^{(k)} - \sigma \cdot s_\ell[\mathbf{a}_0]\|_2 \leq C' 2^{-k}. \quad (\text{C.16})$$

It should be clear that exact recovery is unlikely if \mathbf{x}_0 contains many consecutive nonzero entries: in fact in this situation, even *non-blind* deconvolution fails. Therefore to obtain exact recovery it is necessary to put an upper bound on signal dimension n . Here, we introduce the notation κ_I as an upper bound for number of nonzero entries of \mathbf{x}_0 in a length- p window:

$$\kappa_I := 6 \max\{\theta p, \log n\}, \quad (\text{C.17})$$

where the indexing and addition should be interpreted modulo n . We will denote the support sets of true sparse vector \mathbf{x}_0 and recovered $\mathbf{x}^{(k)}$ in the intermediate k -th steps as

$$I = \text{supp}(\mathbf{x}_0), \quad I^{(k)} = \text{supp}(\mathbf{x}^{(k)}), \quad (\text{C.18})$$

then in the Bernoulli-Gaussian model, with high probability,

$$\max_{\ell} |I \cap ([p] + \ell)| \leq \kappa_I. \quad (\text{C.19})$$

The $\log n$ term reflects the fact that as n becomes enormous (exponential in p) eventually it becomes likely that some length- p window of \mathbf{x}_0 is densely occupied. In our main theorem statement, we preclude this possibility by putting an upper bound on signal length n with respect to window length p and shift coherence μ . We will assume

$$(\mu + 1/p) \cdot \kappa_I^2 < c \quad (\text{C.20})$$

for some numerical constant $c \in (0, 1)$.

Recall that (B.22) in Theorem B.1 provides that

$$\|\bar{\mathbf{a}} - \mathbf{a}_0\|_2 \leq (\mu + 1/p), \quad (\text{C.21})$$

which is sufficiently close to \mathbf{a}_0 as long as (C.19) holds true. Here, we will elaborate this by showing a single iteration of alternating minimization algorithm (C.13)-(C.15) is a contraction mapping for \mathbf{a} toward \mathbf{a}_0 .

To this end, at k -th iteration, write $T = I^{(k)}$, $J = I^{(k+1)}$ and $\boldsymbol{\sigma}^{(k)} = \text{sign}(\mathbf{x}^{(k)})$, then first observe that the solution to the reweighted Lasso problem (C.13) can be written as

$$\mathbf{x}^{(k+1)} = \boldsymbol{\iota}_J \left(\boldsymbol{\iota}_J^* \mathbf{C}_{\mathbf{a}^{(k)}}^* \mathbf{C}_{\mathbf{a}^{(k)}} \boldsymbol{\iota}_J \right)^{-1} \boldsymbol{\iota}_J^* \left(\mathbf{C}_{\mathbf{a}^{(k)}}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0 - \lambda^{(k)} \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}^{(k+1)} \right), \quad (\text{C.22})$$

and the solution to least squares problem (C.14) will be

$$\mathbf{a}^{(k+1)} = (\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}^{(k+1)}}^* \mathbf{C}_{\mathbf{x}^{(k+1)}} \boldsymbol{\iota})^{-1} (\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}^{(k+1)}}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0). \quad (\text{C.23})$$

Here, we are going to illustrate the relationship between $\mathbf{a}^{(k+1)} - \mathbf{a}_0$ and $\mathbf{a}^{(k)} - \mathbf{a}_0$ using simple approximations. First, let us assume that $\mathbf{a}^{(k)} \approx \mathbf{a}_0$, $\mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{a}_0} \approx \mathbf{I}$, and $\mathbf{I} \approx \mathbf{J} \approx \mathbf{T}$. Then (C.22) gives

$$\mathbf{x}^{(k+1)} \approx \mathbf{x}_0, \quad (\text{C.24})$$

$$\begin{aligned} (\mathbf{x}^{(k+1)} - \mathbf{x}_0) &\approx \mathbf{P}_I (\mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0 - \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{a}^{(k)}} \mathbf{x}_0) \\ &\approx \mathbf{P}_I [\mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} (\mathbf{a}_0 - \mathbf{a}^{(k)})], \end{aligned} \quad (\text{C.25})$$

which implies, while assuming $\mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}_0} \approx n\theta \mathbf{I}$, that from (C.23):

$$\begin{aligned} (\mathbf{a}^{(k+1)} - \mathbf{a}_0) &\approx (n\theta)^{-1} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}^{(k+1)}}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 - \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}^{(k+1)}}^* \mathbf{C}_{\mathbf{x}^{(k+1)}} \boldsymbol{\iota} \mathbf{a}_0 \\ &\approx (n\theta)^{-1} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} (\mathbf{x}_0 - \mathbf{x}^{(k+1)}) \\ &\approx (n\theta)^{-1} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_I \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} (\mathbf{a}^{(k)} - \mathbf{a}_0). \end{aligned} \quad (\text{C.26})$$

Now since $\mathbf{C}_{\mathbf{x}_0}^* \mathbf{P}_I \mathbf{C}_{\mathbf{x}_0} \approx n\theta \mathbf{e}_0 \mathbf{e}_0^*$, this suggests that $(n\theta)^{-1} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_I \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}$ approximates a contraction mapping with fixed point \mathbf{a}_0 , as follows:

$$\begin{aligned} (n\theta)^{-1} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_I \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} &\approx \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{e}_0 \mathbf{e}_0^* \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \\ &\approx \mathbf{a}_0 \mathbf{a}_0^*. \end{aligned} \quad (\text{C.27})$$

Hence, if we can ensure all above approximation is sufficiently and increasingly accurate as the iterate proceeds, the alternating minimization essentially is a power method which finds the leading eigenvector of matrix $\mathbf{a}_0 \mathbf{a}_0^*$ —and the solution to this algorithm is apparently \mathbf{a}_0 . Indeed, we prove that the iterates produced by this sequence of operations converge to the ground truth at a linear rate, as long as it is initialized sufficiently nearby:

Theorem C.2 (Linear rate convergence of alternating minimization). *Suppose $\mathbf{y} = \mathbf{a}_0 * \mathbf{x}_0$ where \mathbf{a}_0 is μ -shift coherent and $\mathbf{x}_0 \sim \text{BG}(\theta)$, then there exists some constants C, c, c_μ such that if $(\mu + 1/p) \kappa_I^2 < c_\mu$ and $n > C\theta^{-2} p^2 \log n$, then with probability at least $1 - c/n$, for any starting point $\mathbf{a}^{(0)}$ and $\lambda^{(0)}, I^{(0)}$ such that*

$$\|\mathbf{a}^{(0)} - \mathbf{a}_0\|_2 \leq \mu + 1/p, \quad \lambda^{(0)} = 5\kappa_I(\mu + 1/p), \quad I^{(0)} = \text{supp}(\mathbf{C}_{\mathbf{a}^{(0)}}^* \mathbf{y}), \quad (\text{C.28})$$

and for $k = 1, 2, \dots$:

$$\mathbf{x}^{(k+1)} \leftarrow \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{a}^{(k)} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda^{(k)} \sum_{i \notin I^{(k)}} |\mathbf{x}_i|, \quad (\text{C.29})$$

$$\mathbf{a}^{(k+1)} \leftarrow \mathbf{P}_{\mathbb{S}^{p-1}} [\arg\min_{\mathbf{a}} \frac{1}{2} \|\mathbf{a} * \mathbf{x}^{(k+1)} - \mathbf{y}\|_2^2], \quad (\text{C.30})$$

$$\lambda^{(k+1)} \leftarrow \frac{1}{2} \lambda^{(k)}, \quad I^{(k+1)} \leftarrow \text{supp}(\mathbf{x}^{(k+1)}) \quad (\text{C.31})$$

then

$$\|\mathbf{a}^{(k+1)} - \mathbf{a}_0\|_2 \leq (\mu + 1/p) 2^{-k} \quad (\text{C.32})$$

for every $k = 0, 1, 2, \dots$.

Proof. See [Appendix K.3](#). ■

Remark C.3. *The estimates $\mathbf{x}^{(k)}$ also converges to the ground truth \mathbf{x}_0 at a linear rate.*

D Basic bounds for Bernoulli-Gaussian vectors

In this section, we prove several lemmas pertaining to the sparse random vector $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$.

Lemma D.1 (Support of \mathbf{x}_0). *Let $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ and $I_0 = \text{supp}(\mathbf{x}_0) \subseteq [n]$. Suppose $n > 10\theta^{-1}$, then for any $\varepsilon \in (0, \frac{1}{10})$, with probability at least $1 - \varepsilon$ we have*

$$||I_0| - n\theta| \leq 2\sqrt{n\theta} \log \varepsilon^{-1}. \quad (\text{D.1})$$

And suppose $n \geq C\theta^{-2} \log p$ and θ , then with probability at least $1 - 2/n$, we have

$$\forall t \in [2p] \setminus \{0\}, \quad \frac{1}{2}n\theta^2 \leq |I_0 \cap (I_0 + t)| \leq 2n\theta^2 \quad (\text{D.2})$$

where C is a numerical constant.

Proof. Let $\mathbf{x}_0 = \boldsymbol{\omega} \cdot \mathbf{g} \sim_{\text{i.i.d.}} \text{BG}(\theta)$, notice that the support of the Bernoulli-Gaussian vector \mathbf{x}_0 is almost surely equal to the support of the Bernoulli vector $\boldsymbol{\omega}$. Applying Bernstein inequality [Lemma N.4](#) with $(\sigma^2, R) = (1, 1)$, then if $n\theta > 10$ we have

$$\mathbb{P} \left[\left| \sum_{k \in [n]} \omega_k - n\theta \right| > 2\sqrt{n\theta} \log \varepsilon^{-1} \right] \leq 2 \exp \left(\frac{-4n\theta \log^2 \varepsilon^{-1}}{2n\theta + 4\sqrt{n\theta} \log \varepsilon^{-1}} \right) \leq \varepsilon.$$

For [\(D.2\)](#), let $J_t := I_0 \cap (I_0 + t)$. The cardinality of J_t is an inner product between shifts of $\boldsymbol{\omega}$:

$$|J_t| = \sum_{k \in [n]} \omega_k \omega_{k-t}, \quad (\text{D.3})$$

and define two subset $J_{t_1} \uplus J_{t_2} = J_t$, as follows:

$$\begin{cases} J_{t_1} = J_t \cap \mathcal{K}_1, & \mathcal{K}_1 := [n] \cap \{0, \dots, t-1, 2t, \dots, 3t-1, \dots\} \\ J_{t_2} = J_t \cap \mathcal{K}_2, & \mathcal{K}_2 := [n] \cap \{t, \dots, 2t-1, 3t, \dots, 4t-1, \dots\} \end{cases} \quad (\text{D.4})$$

Here, the size of sets $\mathcal{K}_1, \mathcal{K}_2$ has two-side bounds $0.4n \leq (n - 2p)/2 \leq |\mathcal{K}_2| \leq |\mathcal{K}_1| \leq (n + 2p)/2 \leq 0.6n$, thus the size of sets J_{t_1}, J_{t_2} can be derived using Bernstein inequality [Lemma N.4](#) with $n > C\theta^{-2} \log p$ as

$$\begin{aligned} \mathbb{P} \left[\max_{t \in [2p] \setminus \{0\}} |J_{t_1}| \geq n\theta^2 \right] &= \mathbb{P} \left[\max_{t \in [2p] \setminus \{0\}} \sum_{k \in \mathcal{K}_1} \omega_k \omega_{k-t} \geq n\theta^2 \right] \\ &\leq 2p \cdot \mathbb{P} \left[\sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} \geq n\theta^2 \right] \\ &\leq 2p \cdot \mathbb{P} \left[\sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} - \mathbb{E} \sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} \geq n\theta^2 - 0.6n\theta^2 \right] \\ &\leq 4p \cdot \exp \left(\frac{-(0.4n\theta^2)^2}{2 \cdot 0.6n\theta^2 + 2 \cdot 0.4n\theta^2} \right) = \exp(\log(4p) - 0.08n\theta^2) \\ &\leq 1/n, \end{aligned} \quad (\text{D.5})$$

where the last two inequalities hold with $C > 10^5$. The lower bound can also derived as follows

$$\mathbb{P} \left[\min_{t \in [2p] \setminus \{0\}} |J_{t_1}| \leq n\theta^2/4 \right] = \mathbb{P} \left[\min_{t \in [2p] \setminus \{0\}} \sum_{k \in \mathcal{K}_1} \omega_k \omega_{k-t} \leq n\theta^2/4 \right]$$

$$\begin{aligned}
&\leq 2p \cdot \mathbb{P} \left[\sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} \leq n\theta^2/4 \right] \\
&\leq 2p \cdot \mathbb{P} \left[\sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} - \mathbb{E} \sum_{k \in \mathcal{K}_1} \omega_k \omega_{k+1} \leq n\theta^2/4 - 0.4n\theta^2 \right] \\
&\leq 4p \cdot \exp \left(\frac{-(0.15n\theta^2)^2}{2 \cdot 0.6n\theta^2 + 2 \cdot 0.15n\theta^2} \right) \\
&= \exp(\log(4p) - 0.0015n\theta^2) \leq 1/n.
\end{aligned}$$

The bound for $|J_2|$ can be derived similarly to (D.5)-(D.6). \blacksquare

Lemma D.2 (Norms of \mathbf{x}_0). *Let $\mathbf{x}_0 \sim_{\text{i.i.d.}}$ $\text{BG}(\theta) \in \mathbb{R}^n$. If $n \geq 10\theta^{-1}$, then for any $\varepsilon \in (0, \frac{1}{10})$, with probability at least $1 - \varepsilon$,*

$$\left\| \|\mathbf{x}_0\|_1 - \sqrt{2/\pi}n\theta \right\| \leq 2\sqrt{n\theta} \log \varepsilon^{-1}, \quad \left\| \|\mathbf{x}_0\|_2^2 - n\theta \right\| \leq 3\sqrt{n\theta} \log \varepsilon^{-1} \quad (\text{D.6})$$

Proof. To bound $\|\mathbf{x}_0\|_1$, using Bernstein inequality with $(\sigma^2, R) = (\theta, 1)$ and with $n\theta \geq 10$ we have

$$\mathbb{P} \left[\left| \|\mathbf{x}_0\|_1 - \sqrt{\frac{2}{\pi}}n\theta \right| \geq 2\sqrt{n\theta} \log \varepsilon^{-1} \right] \leq 2 \exp \left(\frac{-4n\theta \log^2 \varepsilon^{-1}}{2n\theta + 4\sqrt{n\theta} \log \varepsilon^{-1}} \right) \leq \varepsilon$$

Similarly for $\|\mathbf{x}_0\|_2^2$, from Gaussian moments [Lemma N.2](#), we know the 2-norm $\sum_{i \in [n]} \mathbb{E} |x_{0i}|^4 = 3n\theta$ and q -norm $\sum_{i \in [n]} \mathbb{E} |x_{0i}|^{2q} \leq (n\theta)(2q-1)!! \leq \frac{1}{2}(3n\theta)2^{q-2}q!$ for $q \geq 3$. Let $(\sigma^2, R) = (3\theta, 2)$ in Bernstein inequality form [Lemma N.4](#), $n\theta \geq 10$ we have

$$\mathbb{P} \left[\left| \|\mathbf{x}_0\|_2^2 - n\theta \right| \geq 3\sqrt{n\theta} \log \varepsilon^{-1} \right] \leq 2 \exp \left(\frac{-9n\theta \log^2 \varepsilon^{-1}}{2(3n\theta) + 12\sqrt{n\theta} \log \varepsilon^{-1}} \right) \leq \varepsilon,$$

completing the proof. \blacksquare

Lemma D.3 (Norms of \mathbf{x}_0 subvectors). *Let $\mathbf{x}_0 \sim_{\text{i.i.d.}}$ $\text{BG}(\theta) \in \mathbb{R}^n$ and $n > 10$, then with probability at least $1 - 3/n$, we have*

$$\max_{\substack{U=[2p]+j \\ j \in [n]}} \|\mathbf{P}_U \mathbf{x}_0\|_2^2 \leq 2p\theta + 6 \left(\sqrt{p\theta} + \log n \right) \quad (\text{D.7})$$

and if \mathbf{a}_0 is μ -shift coherent and there exists a constant c_μ such that both $\theta^2 p < c_\mu$ and $\mu p^2 \theta < c_\mu$, then

$$\max_{\substack{U=[p]+j \\ j \in [n]}} \|\mathbf{P}_U [\mathbf{a}_0 * \mathbf{x}_0]\|_2^2 \leq p\theta + \log n. \quad (\text{D.8})$$

Proof. Use Bernstein inequality with $(\sigma^2, R) = (3\theta, 2)$ and $t = \max \{ \sqrt{p\theta}, \log n \}$, with union bound we obtain:

$$\begin{aligned}
\mathbb{P} \left[\max_{\substack{U=[2p]+j \\ j \in [n]}} \|\mathbf{P}_U \mathbf{x}_0\|_2^2 \geq 2p\theta + 6 \left(\sqrt{p\theta} + \log n \right) \right] &\leq 2n \exp \left(-\frac{36 \left(\sqrt{p\theta} + \log n \right)^2}{6p\theta + 12 \left(\sqrt{p\theta} + \log n \right)} \right) \\
&\leq 2 \exp \left(\log n - \frac{36t^2}{6t^2 + 12t} \right) \leq \frac{2}{n}. \quad (\text{D.9})
\end{aligned}$$

For the second inequality, first we know calculate the expectation

$$\begin{aligned}
\mathbb{E} \|\mathbf{P}_U [\mathbf{a}_0 * \mathbf{x}_0]\|_2^2 &= \mathbb{E} [\mathbf{x}_0^* \mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_U \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0] \\
&= \theta \cdot \text{tr} (\mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_U \mathbf{C}_{\mathbf{a}_0}) \|\mathbf{a}_0\|_2^2 + \theta \cdot \sum_{i=1}^{p-1} \|\boldsymbol{\iota}^* s_i[\mathbf{a}_0]\|_2^2 \\
&= p\theta.
\end{aligned} \tag{D.10}$$

Then apply Henson Wright inequality [Lemma N.6](#) with $\|\mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_U \mathbf{C}_{\mathbf{a}_0}\|_F^2 = \|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_F^2 \leq p(1 + \mu p)$ and also $\|\mathbf{C}_{\mathbf{a}_0}^* \mathbf{P}_U \mathbf{C}_{\mathbf{a}_0}\|_2 = \|\mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_2^2 = 1 + \mu p$, we can derive

$$\begin{aligned}
\mathbb{P} \left[\max_{\substack{U=[p]+j \\ j \in [n]}} \|\mathbf{P}_U [\mathbf{a}_0 * \mathbf{x}_0]\|_2^2 \geq p\theta + \log n \right] &\leq n \exp \left(- \min \left\{ \frac{\log^2 n}{64\theta^2 p(1 + \mu p)}, \frac{\log n}{8\sqrt{2}\theta(1 + \mu p)} \right\} \right) \\
&\leq \exp \left(\log n - \min \left\{ \frac{\log^2 n}{128c_\mu}, \frac{\log n}{32c_\mu} \right\} \right) \leq \frac{1}{n}
\end{aligned} \tag{D.11}$$

when $c_\mu < \frac{1}{300}$. ■

Lemma D.4 (Inner product between shifted \mathbf{x}_0). *Let $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^n$. There exists a numerical constant C such that if $n > C\theta^{-2} \log p$ and $p\theta \log^2 \theta^{-1} > 1$, with probability at least $1 - 4/n$, the following two statements hold simultaneously:*

$$\max_{i \neq j \in [2p]} \langle s_i[\mathbf{x}_0], s_j[\mathbf{x}_0] \rangle \leq 6\sqrt{n\theta^2 \log n}; \tag{D.12}$$

and for $\mathbf{x}_i = |\mathbf{x}_{0,i}| \in \mathbb{R}_+^n$ the vector of magnitudes of \mathbf{x}_0 ,

$$\max_{i \neq j \in [2p]} \langle s_i[\mathbf{x}], s_j[\mathbf{x}] \rangle \leq 4n\theta^2. \tag{D.13}$$

Proof. We will start from proving (D.13). Write $\mathbf{x} = |\mathbf{g}| \circ \boldsymbol{\omega}$ where $\mathbf{g} / \boldsymbol{\omega}$ are Gaussian/Bernoulli random vectors respectively. Let I_0 denote the support of $\boldsymbol{\omega}$ and $t = |j - i|$ with $0 < t < p$. Then (D.13) can be written as summation of Gaussian r.v.s. on intersection of support set between shifts:

$$\langle s_i[\mathbf{x}], s_j[\mathbf{x}] \rangle = \sum_{k \in I_0 \cap (I_0 + t)} |\mathbf{g}_k| |\mathbf{g}_{k-t}| \tag{D.14}$$

Define $J_t := I_0 \cap (I_0 + t) = J_{t1} \uplus J_{t2}$ same as (D.4). Notice that both $\sum_{k \in J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k-t}|$ and $\sum_{k \in J_{t2}} |\mathbf{g}_k| |\mathbf{g}_{k-t}|$ are sum of independent r.v.s.. We are left to consider the upper bound of $\sum_{j \in J_{ti}} |\mathbf{g}_j| |\mathbf{g}'_j|$ where \mathbf{g}, \mathbf{g}' are independent Gaussian vectors. We condition on the following event

$$\mathcal{E}_J := \{ \forall t \in [2p] \setminus \{0\}, n\theta^2/4 \leq |J_{t1}|, |J_{t2}| \leq n\theta^2 \}, \tag{D.15}$$

which holds w.p. at least $1 - 2/n$ from [Lemma D.1](#). Since $\sum_{j \in J_{t1}} |\mathbf{g}_j| |\mathbf{g}'_j| \leq \|\mathbf{g}_{J_{t1}}\|_2 \|\mathbf{g}'_{J_{t1}}\|_2$, we use Gaussian concentration [Lemma N.3](#) and union bound to obtain

$$\begin{aligned}
\mathbb{P} \left[\max_{t \in [2p] \setminus \{0\}} \sum_{j \in J_{t1}} |\mathbf{g}_j \mathbf{g}'_j| > 2|J_{t1}| \right] &\leq 2p \cdot \mathbb{P} [\|\mathbf{g}_{J_{t1}}\|_2 \|\mathbf{g}'_{J_{t1}}\|_2 - \mathbb{E} \|\mathbf{g}_{J_{t1}}\|_2 \|\mathbf{g}'_{J_{t1}}\|_2 > |J_{t1}|] \\
&\leq 4p \cdot \mathbb{P} [\|\mathbf{g}_{J_{t1}}\|_2 - \mathbb{E} \|\mathbf{g}_{J_{t1}}\|_2 > \sqrt{|J_{t1}|}/3] \\
&\leq 4p \exp(-(|J_{t1}|/9)/2) \leq 4p \exp(-n\theta^2/72) \leq 1/n
\end{aligned} \tag{D.16}$$

where the last inequality is derived simply via assuming $n = C\theta^{-2} \log p$ for some $C > 10^4$, such that

$$\begin{aligned} C > 400 * (4C)^{1/5} &\implies C \log p > 400 \log((4C)^{1/5} p) \\ &\implies C \log p > 72 \log(4Cp^5) > 72 \log(4Cp^2 \log^3 p) \\ &\implies n\theta^2 > 72 \log(p \cdot 4C\theta^{-2} \log p) = 72 \log(4np). \end{aligned}$$

Likewise for sum on set J_{t2} , we collect all above result and conclude for every $i \neq j \in [2p]$,

$$\langle s_i[\mathbf{x}], s_j[\mathbf{x}] \rangle = \sum_{k \in J_{t1}} |\mathbf{g}_k| |\mathbf{g}'_{k-t}| + \sum_{k \in J_{t2}} |\mathbf{g}_k| |\mathbf{g}'_{k-t}| \leq 2(|J_{t1}| + |J_{t2}|) \leq 4n\theta^2. \quad (\text{D.17})$$

For (D.12) similarly condition on event \mathcal{E}_J , using Bernstein inequality [Lemma N.4](#) with $(\sigma^2, R) = (1, 1)$:

$$\begin{aligned} \mathbb{P} \left[\max_{t \in [2p] \setminus \{0\}} \left| \sum_{j \in J_{t1}} \mathbf{g}_j \mathbf{g}'_j \right| > 3\sqrt{n\theta^2 \log n} \right] &\leq p \cdot \exp \left(\frac{-9n\theta^2 \log n}{2|J_{t1}| + 6\sqrt{n\theta^2 \log n}} \right) \\ &\leq p \cdot \exp \left(\frac{-9n\theta^2 \log n}{3n\theta^2} \right) \leq \frac{1}{n} \end{aligned} \quad (\text{D.18})$$

thus for every $i \neq j \in [2p]$,

$$|\langle s_i[\mathbf{x}_0], s_j[\mathbf{s}_0] \rangle| \leq \left| \sum_{k \in J_{t1}} \mathbf{g}_k \mathbf{g}'_{k-t} \right| + \left| \sum_{k \in J_{t2}} \mathbf{g}_k \mathbf{g}'_{k-t} \right| \leq 6\sqrt{n\theta^2 \log n}. \quad (\text{D.19})$$

Finally, both (D.17),(D.19) holds simultaneously with probability at least

$$1 - 2/n - 1/n - 1/n = 1 - 4/n \quad (\text{D.20})$$

■

Lemma D.5 (Convolution of \mathbf{x}_0). *Given $\mathbf{y} = \mathbf{x}_0 * \mathbf{a}_0$ where $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^n$ and $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ is μ -shift coherent. Suppose $n \geq C\theta^{-2} \log p$ for some numerical constant $C > 0$, with probability at least $1 - 7/n$, we have the following two statement simultaneously hold:*

$$\|\mathbf{C}_{\mathbf{y}} \boldsymbol{\iota}\|_2^2 \leq 3(1 + \mu p)n\theta \quad (\text{D.21})$$

and for all $J \subseteq [n]$,

$$\|\mathbf{P}_J \mathbf{C}_{\mathbf{y}} \boldsymbol{\iota}\|_2^2 \leq 14|J|(1 + \mu p)(p\theta + \log n) \quad (\text{D.22})$$

Proof. Given any $\mathbf{a} \in \mathbb{S}^{p-1}$, write $\boldsymbol{\beta} = \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \mathbf{a}$ where $|\boldsymbol{\beta}| \leq 2p$. Apply $\|\mathbf{x}_0\|_2^2 \leq 2n\theta$ from [Lemma D.2](#) by choosing $\varepsilon = 1/n$, also $|\langle s_i[\mathbf{x}_0], s_j[\mathbf{x}_0] \rangle| \leq 6\sqrt{n\theta^2 \log n}$ from [Lemma D.4](#) we get:

$$\begin{aligned} \|\mathbf{C}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a}\|_2^2 &= \|\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\beta}\|_2^2 \leq \|\boldsymbol{\beta}\|_2^2 \|\mathbf{x}_0\|_2^2 + \sum_{i \neq j \in [\pm p]} |\beta_i \beta_j| |\langle s_i[\mathbf{x}_0], s_j[\mathbf{x}_0] \rangle| \\ &\leq \|\boldsymbol{\beta}\|_2^2 \|\mathbf{x}_0\|_2^2 + \|\boldsymbol{\beta}\|_1 \max_{i \neq j \in [\pm p]} |\langle s_i[\mathbf{x}_0], s_j[\mathbf{x}_0] \rangle| \\ &\leq \|\boldsymbol{\beta}\|_2^2 \cdot 2n\theta + p \|\boldsymbol{\beta}\|_2^2 \cdot 6\sqrt{n\theta^2 \log n} \leq 3 \|\boldsymbol{\beta}\|_2^2 n\theta \end{aligned}$$

where $n = C\theta^{-2} \log p$ with $C \geq 10^4$, and the statement holds with probability at least $1 - 5/n$.

For the bound of $\|\mathbf{P}_J \mathbf{C}_y \boldsymbol{\iota} \boldsymbol{\alpha}\|_2^2$. Simply apply [Lemma D.3](#) and utilize norm bound of $\|\boldsymbol{\beta}\|_2^2$, with probability at least $1 - 2/n$ we have:

$$\|\mathbf{P}_J \mathbf{C}_y \boldsymbol{\iota} \boldsymbol{\alpha}\|_2^2 = \sum_{i \in J} |\langle s_i[\mathbf{x}_0], \boldsymbol{\beta} \rangle|^2 \leq |J| \max_{\substack{U=[2p]+j \\ j \in [n]}} \|\mathbf{P}_U \mathbf{x}_0\|_2^2 \|\boldsymbol{\beta}\|_2^2 \leq |J| \cdot 14(p\theta + \log n) \cdot \|\boldsymbol{\beta}\|_2^2$$

Finally apply [Lemma E.4](#) and Gershgorin disc theorem obtain

$$\|\boldsymbol{\beta}\|_2^2 = \|\mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \boldsymbol{\alpha}\|_2^2 \leq \|\mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}\|_2^2 = \sigma_{\max}(\mathbf{M}) \leq 1 + \mu p. \quad (\text{D.23})$$

Remark D.6. When \mathbf{a}_0 is a basis vector \mathbf{e}_0 , the result of [Lemma D.5](#) gives upper bound of $\|\mathbf{C}_{\mathbf{x}_0}\|_2 < 3n\theta$, whose lower bound can be derived similarly with $\|\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}\|_2 \geq \frac{2}{3}n\theta$ ■

E Vectors in shift space

In this section, we will establish a number of properties of the coefficient vectors $\boldsymbol{\alpha}$ and correlation vector $\boldsymbol{\beta}$. Generally speaking, when \mathbf{a} is close to the subspace \mathcal{S}_τ , then both vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$ have most of their energy concentrated on the entries τ . In this section, we derive upper bounds on $\boldsymbol{\alpha}_{\tau^c}$ and $\boldsymbol{\beta}_{\tau^c}$ under various assumptions.

In particular, we will introduce a relationship between the sparsity rate θ , coherence μ and size $|\tau|$, which we term the sparsity-coherence condition. In [Lemma E.2](#) we prove that measuring the distance from \mathbf{a} to subspace \mathcal{S}_τ in terms of $\|\boldsymbol{\alpha}_{\tau^c}\|_2$ gives a seminorm. We then use this distance to characterize a region $\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ around the subspace \mathcal{S}_τ . Later, in [Lemma E.4](#) we illustrate the relationship between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}$. Finally in [Lemma E.5](#) and [Corollary E.6](#), controls the magnitude of $\boldsymbol{\alpha}_{\tau^c}$ and $\boldsymbol{\beta}_{\tau^c}$ near \mathcal{S}_τ .

Definition E.1 (Sparsity-coherence condition). Let $\mathbf{a}_0 \in \mathbb{S}^{p_0-1}$ with shift coherence μ . We say that $(\mathbf{a}_0, \theta, |\tau|)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$ with constant c_μ , if

$$\theta \in \left[\frac{1}{p}, \frac{c_\mu}{4 \max\{|\tau|, \sqrt{p}\}} \right] \cdot \frac{1}{\log^2 \theta^{-1}}, \quad \mu \cdot \max\{|\tau|^2, p^2 \theta^2\} \cdot \log^2 \theta^{-1} \leq \frac{c_\mu}{4}, \quad (\text{E.1})$$

where $p = 3p_0 - 2$.

Lemma E.2 (d_α is a seminorm). For every solution subspace \mathcal{S}_τ , the function $d_\alpha(\cdot, \mathcal{S}_\tau) : \mathbb{R}^p \rightarrow \mathbb{R}_+$ defined as

$$d_\alpha(\mathbf{a}, \mathcal{S}_\tau) = \inf \{\|\boldsymbol{\alpha}_{\tau^c}\|_2 \mid \mathbf{a} = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}\}. \quad (\text{E.2})$$

is a seminorm, and for all $\mathbf{a} \in \mathcal{S}_\tau$, $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) = 0$.

Proof. It is immediate from definition that $d(\cdot, \mathcal{S}_\tau)$ is nonnegative and $\mathcal{S}_\tau \subseteq \{\mathbf{a} : d_\alpha(\mathbf{a}, \mathcal{S}_\tau) = 0\}$. Subadditivity can be shown from simple norm inequalities and our definition of d_α , for all $\mathbf{a}_1, \mathbf{a}_2$ we have

$$\begin{aligned} d_\alpha(\mathbf{a}_1 + \mathbf{a}_2, \mathcal{S}_\tau) &= \inf \{\|\boldsymbol{\alpha}_{\tau^c}\|_2 \mid \mathbf{a}_1 + \mathbf{a}_2 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}\} \\ &= \inf \{\|\boldsymbol{\alpha}_{1\tau^c} + \boldsymbol{\alpha}_{2\tau^c}\|_2 \mid \mathbf{a}_1 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_1, \quad \mathbf{a}_2 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_2\} \\ &\leq \inf \{\|\boldsymbol{\alpha}_{1\tau^c}\|_2 + \|\boldsymbol{\alpha}_{2\tau^c}\|_2 \mid \mathbf{a}_1 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_1, \quad \mathbf{a}_2 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_2\} \\ &= \inf \{\|\boldsymbol{\alpha}_{1\tau^c}\|_2 \mid \mathbf{a}_1 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_1\} + \inf \{\|\boldsymbol{\alpha}_{2\tau^c}\|_2 \mid \mathbf{a}_2 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}_2\} \\ &= d_\alpha(\mathbf{a}_1, \mathcal{S}_\tau) + d_\alpha(\mathbf{a}_2, \mathcal{S}_\tau). \end{aligned}$$

Similarly the absolute homogeneity, for any $c \in \mathbb{R}$:

$$d_\alpha(c \cdot \mathbf{a}, \mathcal{S}_\tau) = \inf \{\|\boldsymbol{\alpha}'_{\tau^c}\|_2 \mid c \cdot \mathbf{a} = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}'\} = \inf \{\|c \cdot \boldsymbol{\alpha}_{\tau^c}\|_2 \mid \mathbf{a} = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}\}$$

$$= |c| \cdot \inf \{ \|\boldsymbol{\alpha}_{\tau^c}\|_2 \mid \boldsymbol{a} = \boldsymbol{\iota}^* \boldsymbol{C}_{\mathbf{a}_0} \boldsymbol{\alpha} \} = |c| \cdot d_\alpha(\boldsymbol{a}, \mathcal{S}_\tau),$$

which completes the proof that d_α is a seminorm. \blacksquare

Definition E.3 (Widened subspace). For subspace \mathcal{S}_τ let

$$\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu)) := \{ \boldsymbol{a} \in \mathbb{S}^{p-1} \mid d_\alpha(\boldsymbol{a}, \mathcal{S}_\tau) \leq \gamma \} \quad (\text{E.3})$$

denote its widening by γ , in the seminorm d_α .

Our analysis works with a specific choice of width $\gamma(c_\mu)$, which depends on the problem parameters $\mathbf{a}_0, \theta, |\tau|$ and a constant c_μ , via

$$\gamma(c_\mu) = \frac{c_\mu}{4 \log^2 \theta^{-1}} \min \left\{ \frac{1}{\sqrt{|\tau|}}, \frac{1}{\sqrt{\mu p}}, \frac{1}{\mu p \sqrt{\theta} |\tau|} \right\} \quad (\text{E.4})$$

Lemma E.4 (Properties of $\boldsymbol{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}^* \boldsymbol{C}_{\mathbf{a}_0}$). Let $\boldsymbol{M} = \boldsymbol{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}^* \boldsymbol{C}_{\mathbf{a}_0}$, with $\mathbf{a}_0 \in \mathbb{S}^{p_0-1}$ μ -shift coherent. The diagonal entries of \boldsymbol{M} satisfy

$$\begin{cases} \boldsymbol{M}_{ii} = 1 & i \in [-p_0 + 1, p_0 - 1] = [\pm p_0], \\ 0 \leq \boldsymbol{M}_{ii} \leq 1 & i \in [-2p_0 + 2, -p_0] \cup [p_0, 2p_0 - 2], \\ \boldsymbol{M}_{ii} = 0 & \text{otherwise,} \end{cases} \quad (\text{E.5})$$

and the off-diagonal entries satisfy

$$\begin{cases} |\boldsymbol{M}_{ij}| \leq \mu & 0 < |i - j| < p_0, \quad \{i \in [-p_0 + 1, p_0 - 1]\} \cup \{j \in [-p_0 + 1, p_0 - 1]\} \\ |\boldsymbol{M}_{ij}| < 1 & \{i, j \in [-2p_0 + 2, -p_0]\} \cup \{i, j \in [p_0, 2p_0 - 2]\} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{E.6})$$

Furthermore, let $\tau \subset [\pm p_0]$, and $\tau^c = [\pm 2p_0 - 1] \setminus \tau$. The singular values of submatrix $\boldsymbol{\iota}_\tau^* \boldsymbol{M} \boldsymbol{\iota}_\tau$ can be bounded as:

$$\begin{cases} 1 - \mu |\tau| \leq \sigma_{\min}(\boldsymbol{\iota}_\tau^* \boldsymbol{M} \boldsymbol{\iota}_\tau) \leq \sigma_{\max}(\boldsymbol{\iota}_\tau^* \boldsymbol{M} \boldsymbol{\iota}_\tau) \leq 1 + \mu |\tau| \\ \sigma_{\max}(\boldsymbol{\iota}_{\tau^c}^* \boldsymbol{M} \boldsymbol{\iota}_{\tau^c}) \leq \mu \sqrt{p |\tau|} \\ \sigma_{\max}(\boldsymbol{\iota}_{\tau^c}^* \boldsymbol{M} \boldsymbol{\iota}_{\tau^c}) \leq 1 + \mu p \end{cases} \quad (\text{E.7})$$

Proof. Recall the definition of $\boldsymbol{\iota}$, which selects the entries $\{-p_0 + 1, \dots, 2p_0 - 2\}$. The entrywise properties of \boldsymbol{M} can be derived by carefully counting the entries of the shifted support. The submatrix \boldsymbol{M} on support $\{-2p_0 + 2, \dots, 2p_0 - 2\}$ has an upper bound to be characterized as follows:

$$\left| \boldsymbol{\iota}_{[\pm 2p_0 - 1]}^* \boldsymbol{M} \boldsymbol{\iota}_{[\pm 2p_0 - 1]} \right| \leq \begin{bmatrix} \boldsymbol{J} & \mu \cdot \mathbf{1} & \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} & \mathbf{0} & \mathbf{0} \\ \mu \cdot \mathbf{1} & \boldsymbol{I} + \mu \cdot \mathbf{1}_o & \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} & \mu \cdot \mathbf{1} & \mathbf{0} \\ [0 \cdots 0] & [\mu \cdots \mu] & 1 & [\mu \cdots \mu] & [0 \cdots 0] \\ \mathbf{0} & \mu \cdot \mathbf{1} & \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} & \boldsymbol{I} + \mu \cdot \mathbf{1}_o & \mu \cdot \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} & \mu \cdot \mathbf{1} & \boldsymbol{J} \end{bmatrix}. \quad (\text{E.8})$$

Here, the center row/column vector is indexed at 0, the matrices \mathbf{J} , \mathbf{I} , $\mathbf{1}$ and $\mathbf{1}_o$ are square and of size $(p_0 - 1)^2$. Among which, \mathbf{I} is the identity matrix, $\mathbf{1}$ is the ones matrix whereas $\mathbf{1}_o$ has all off diagonal entries equal 1. Also $|\mathbf{J}|$ has property $|\mathbf{J}_{ij}| < 1$ for all i, j .

As for the singular values, notice that the first and second inequalities consider submatrix not containing \mathbf{J} since $\tau \subseteq [\pm p_0]$; thus the first inequality can be derived with Gershgorin disc theorem directly, and the second inequality with the upper bound with its Frobenius norm:

$$\sigma_{\max}(\iota_{\tau^c}^* \mathbf{M} \iota_{\tau}) \leq \mu \sqrt{(2p_0 - 1) |\tau|} < \mu \sqrt{p |\tau|}. \quad (\text{E.9})$$

Finally by recalling $p = 3p_0 - 2 > 2p_0 - 1$. The last inequality is direct from bound of $\iota^* \mathbf{C}_{a_0}$:

$$\sigma_{\max}(\iota_{\tau^c}^* \mathbf{M} \iota_{\tau^c}) \leq \|\mathbf{C}_{a_0}^* \iota^* \mathbf{C}_{a_0}\|_2 = \|\iota^* \mathbf{C}_{a_0} \mathbf{C}_{a_0}^* \iota\|_2 = \|\iota^* \mathbf{C}_{a_0}^* \mathbf{C}_{a_0} \iota\|_2 \leq 1 + \mu p \quad (\text{E.10})$$

where the third equality is derived via commutativity of convolution. \blacksquare

Lemma E.5 (Shift space vectors in widened subspace). *Let $(a_0, \theta, |\tau|)$ satisfy the sparsity-coherence condition $\text{SCC}(c_\mu)$. Then for every $\mathbf{a} \in \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$, every α satisfying $\mathbf{a} = \iota^* \mathbf{C}_{a_0} \alpha$ and $\|\alpha_{\tau^c}\|_2 \leq \gamma(c_\mu)$ has*

$$\|\|\alpha_\tau\|_2 - 1\| \leq c_\mu; \quad (\text{E.11})$$

moreover, $\beta = \mathbf{C}_{a_0}^* \iota \mathbf{a}$ satisfies

$$1 - 3c_\mu \leq \|\beta_\tau\|_2^2 \leq 1 + \frac{c_\mu}{|\tau| \log^2 \theta^{-1}} \quad (\text{E.12})$$

$$\|\beta_{\tau^c}\|_\infty \leq \frac{c_\mu}{\sqrt{|\tau|} \log^2 \theta^{-1}} \quad (\text{E.13})$$

$$\|\beta_{\tau^c}\|_2 \leq \frac{c_\mu}{|\tau| \theta \log \theta^{-1}} \min\{\sqrt{\theta}, \gamma(c_\mu)\}. \quad (\text{E.14})$$

Proof. Write $-1/\log \theta = \theta_{\log}$ and $\gamma = \gamma(c_\mu)$ for convenience. First, by using bounds on γ in (E.4) and $\mu |\tau| < 1$ we obtain:

$$\begin{cases} \gamma \cdot \sqrt{1 + \mu p} \leq \gamma (1 + \sqrt{\mu p}) \leq c_\mu \theta_{\log}^2 / 2 \\ \gamma \cdot \sqrt{1 + \mu^2 p} \leq \gamma (1 + \sqrt{\mu^2 p}) \leq \frac{c_\mu \theta_{\log}^2}{4} \left(\frac{1}{\sqrt{|\tau|}} + \sqrt{\mu} \right) \leq \frac{c_\mu \theta_{\log}^2}{2\sqrt{|\tau|}} \\ \gamma \cdot \mu \sqrt{p |\tau|} \leq \gamma \cdot \sqrt{\mu p} \cdot \sqrt{\mu |\tau|} \leq c_\mu \theta_{\log}^2 / 4 \end{cases} \quad (\text{E.15})$$

Let $\mathbf{a} = \iota^* \mathbf{C}_{a_0} \alpha$ with $\|\alpha_{\tau^c}\|_2 < \gamma$. Utilize properties of $\iota^* \mathbf{C}_{a_0}$ from Lemma E.4 and $\mu |\tau| < c_\mu/4$ and (E.15), we have:

$$\begin{aligned} \|\alpha_\tau\|_2 &\geq \|\iota^* \mathbf{C}_{a_0} \iota_\tau\|_2^{-1} (\|\mathbf{a}\|_2 - \|\iota^* \mathbf{C}_{a_0} \alpha_{\tau^c}\|_2) \\ &\geq \|\iota^* \mathbf{C}_{a_0} \iota_\tau\|_2^{-1} (1 - \|\iota^* \mathbf{C}_{a_0}\|_2 \|\alpha_{\tau^c}\|_2) \\ &\geq \frac{1}{\sqrt{1 + \mu |\tau|}} (1 - \gamma \cdot \sqrt{1 + \mu p}) \geq \frac{1 - c_\mu/2}{\sqrt{1 + c_\mu/4}} \geq 1 - c_\mu, \end{aligned} \quad (\text{E.16})$$

and similarly, the upper bound can be derived as:

$$\begin{aligned} \|\alpha_\tau\|_2 &\leq \sigma_{\min}^{-1}(\iota^* \mathbf{C}_{a_0} \iota_\tau) (\|\mathbf{a}\|_2 + \|\iota^* \mathbf{C}_{a_0} \alpha_{\tau^c}\|_2) \\ &\leq \sigma_{\min}^{-1}(\iota^* \mathbf{C}_{a_0} \iota_\tau) (1 + \|\iota^* \mathbf{C}_{a_0}\|_2 \|\alpha_{\tau^c}\|_2) \\ &\leq \frac{1}{\sqrt{1 - \mu |\tau|}} (1 + \gamma \cdot \sqrt{1 + \mu p}) \leq \frac{1 + c_\mu/2}{\sqrt{1 - c_\mu/4}} \leq 1 + c_\mu. \end{aligned} \quad (\text{E.17})$$

The bound of $\|\beta_{\tau}\|_2^2$ can be simply obtained using $\mu|\tau| < c_{\mu}/4$ and γ bound from (E.15) as:

$$\|\beta_{\tau}\|_2^2 \leq \sigma_{\max}^2(\mathbf{u}_{\tau}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{u}_{\tau}) \leq 1 + \mu|\tau| \leq 1 + \frac{c_{\mu}\theta_{\log}^2}{|\tau|} \quad (\text{E.18})$$

$$\begin{aligned} \|\beta_{\tau}\|_2^2 &\geq (\sigma_{\min}(\mathbf{u}_{\tau}^* \mathbf{M} \mathbf{u}_{\tau}) \|\alpha_{\tau}\|_2 - \sigma_{\max}(\mathbf{u}_{\tau}^* \mathbf{M} \mathbf{u}_{\tau^c}) \|\alpha_{\tau^c}\|_2)^2 \\ &\geq \left((1 - \mu|\tau|)(1 - c_{\mu}) - \mu\sqrt{p|\tau|} \cdot \gamma \right)^2 \geq 1 - 3c_{\mu}. \end{aligned} \quad (\text{E.19})$$

As for the upper bound of and $\|\beta_{\tau^c}\|_{\infty}$, follow from (E.15), we have:

$$\begin{aligned} \|\beta_{\tau^c}\|_{\infty} &\leq \|\mathbf{u}_{\tau^c}^* \mathbf{M} \alpha_{\tau}\|_{\infty} + \|\mathbf{u}_{\tau^c}^* \mathbf{M} \alpha_{\tau^c}\|_{\infty} \leq \mu\sqrt{|\tau|} \|\alpha_{\tau}\|_2 + \sqrt{1 + \mu^2 p} \|\alpha_{\tau^c}\|_2 \\ &\leq \frac{c_{\mu}\theta_{\log}^2(1 + c_{\mu})}{4|\tau|} + \gamma \cdot \sqrt{1 + \mu^2 p} \leq \frac{c_{\mu}\theta_{\log}^2}{\sqrt{|\tau|}}; \end{aligned} \quad (\text{E.20})$$

the bound for $\|\beta_{\tau^c}\|_2$ requires two inequalities, we know

$$\|\beta_{\tau^c}\|_2 \leq \|\mathbf{u}_{\tau^c}^* \mathbf{M} \alpha_{\tau}\|_2 + \|\mathbf{u}_{\tau^c}^* \mathbf{M} \alpha_{\tau^c}\|_2 \leq \mu\sqrt{p|\tau|} \|\alpha_{\tau}\|_2 + (1 + \mu p) \|\alpha_{\tau^c}\|_2, \quad (\text{E.21})$$

for the first inequality, use $(\mu|\tau|^2)^{3/4} (\mu p^2 \theta^2)^{1/4} = \mu\sqrt{p\theta} |\tau|^{3/2} < c_{\mu}\theta_{\log}^2/4$, definition of γ and $\theta|\tau| \leq c_{\mu}\theta_{\log}^2/4$ we have:

$$\begin{aligned} (\text{E.21}) &\leq \frac{\mu\sqrt{p\theta} |\tau|^{3/2}}{\sqrt{\theta} |\tau|} (1 + c_{\mu}) + \frac{\sqrt{\theta} |\tau| \cdot \sqrt{|\tau|} \gamma}{\sqrt{\theta} |\tau|} + \frac{\mu p \sqrt{\theta} |\tau| \gamma}{\sqrt{\theta} |\tau|} \\ &\leq \frac{2c_{\mu}\theta_{\log}^2 + c_{\mu}\theta_{\log}^3 + c_{\mu}\theta_{\log}^2}{4\sqrt{\theta} |\tau|} \leq \frac{c_{\mu}\theta_{\log}^2}{\sqrt{\theta} |\tau|}, \end{aligned} \quad (\text{E.22})$$

and similarly for the second inequality, use both conditions of μ , we have:

$$\begin{aligned} (\text{E.21}) &\leq \frac{\gamma}{\theta|\tau|} \cdot \frac{\mu\sqrt{p\theta} |\tau|^{3/2}}{\gamma} (1 + c_{\mu}) + \gamma + \mu p \gamma \\ &\leq \frac{\gamma}{\theta|\tau|} \cdot \frac{4\mu\sqrt{p\theta} |\tau|^{3/2}}{c_{\mu}\theta_{\log}^2} \cdot \max\left\{ \sqrt{|\tau|}, \sqrt{\mu p}, \mu p \sqrt{\theta} |\tau| \right\} \\ &\quad + \frac{\gamma}{\theta|\tau|} \cdot \theta |\tau| + \frac{\gamma}{\theta|\tau|} \cdot \mu p \theta |\tau| \\ &\leq \frac{\gamma}{\theta|\tau|} \cdot \left(\frac{4}{c_{\mu}\theta_{\log}^2} \cdot \max\left\{ \mu|\tau|^2 \cdot \sqrt{p\theta}, \mu(p\theta) |\tau| \cdot \sqrt{\mu|\tau|}, \right. \right. \\ &\quad \left. \left. \mu\sqrt{p\theta} |\tau|^{3/2} \cdot \mu p \theta |\tau| \right\} + \frac{c_{\mu}\theta_{\log}^2}{4} + \frac{c_{\mu}\theta_{\log}^2}{4} \right) \\ &\leq \frac{\gamma}{\theta|\tau|} \left(\frac{c_{\mu}\theta_{\log}}{4} + \frac{c_{\mu}\theta_{\log}^2}{4} + \frac{c_{\mu}\theta_{\log}^2}{4} \right) \leq \frac{c_{\mu}\theta_{\log}\gamma}{\theta|\tau|}, \end{aligned} \quad (\text{E.23})$$

which completes the proof. \blacksquare

Corollary E.6 ($|\langle \beta_{\tau^c}, \mathbf{x}_{0, \tau^c} \rangle|$ is small). *Given $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n and $|\tau|, c_{\mu}$ such that $(\mathbf{a}_0, \theta, |\tau|)$ satisfies the sparsity-coherence condition $\text{SCC}(c_{\mu})$. Write $\lambda = c_{\lambda}/\sqrt{|\tau|}$ with some $c_{\lambda} \geq 1/5$, then if $c_{\mu} \leq \frac{c_{\lambda}}{25}$,*

$$\mathbb{P} \left[\left| \sum_{i \in \tau^c} \beta_i \mathbf{x}_{0i} \right| > \frac{\lambda}{10} \right] \leq 2\theta, \quad \mathbb{P} \left[\left| \sum_i \beta_i \mathbf{x}_{0i} \right| > \frac{\lambda}{10} \right] \leq \theta |\tau| + 2\theta. \quad (\text{E.24})$$

Proof. We bound tail probability of the first result with Gaussian moments [Lemma N.2](#) and Bernstein inequality [Lemma N.4](#). Via Hölder's inequality, $\sum_{i \in \tau^c} \mathbb{E}(\beta_i x_i)^q = \mathbb{E} x_0^q \|\beta_{\tau^c}\|_q^q \leq \theta(q-1)!! \|\beta_{\tau^c}\|_2^2 \|\beta_{\tau^c}\|_\infty^{q-2}$, thus

$$\mathbb{P} \left[\left| \sum_{i \in \tau^c} \beta_i x_{0i} \right| > \lambda/10 \right] \leq 2 \exp \left(\frac{-(\lambda/10)^2}{2\theta \|\beta_{\tau^c}\|_2^2 + 2(\lambda/10) \|\beta_{\tau^c}\|_\infty} \right) \quad (\text{E.25})$$

Write $\theta_{\log} = -\frac{1}{\log \theta}$, [Lemma E.5](#) implies when $c_\mu \leq \frac{c_\lambda}{25}$, we have $\theta \|\beta_{\tau^c}\|_2^2 \leq \frac{c_\mu^2 \theta_{\log}^2}{|\tau|^2} \leq \frac{\theta_{\log} \lambda^2}{625}$ and $\|\beta_{\tau^c}\|_\infty \leq \frac{c_\mu \theta_{\log}}{\sqrt{|\tau|}} \leq \frac{\theta_{\log} \lambda}{25}$, therefore,

$$\begin{aligned} (\text{E.25}) &\leq 2 \exp \left(\frac{-\lambda^2/100}{2\theta_{\log} \lambda^2/625 + 2(\theta_{\log} \lambda/25) \cdot (\lambda/10)} \right) \\ &\leq 2 \exp(\log \theta) \leq 2\theta \end{aligned} \quad (\text{E.26})$$

The second tail bound is straight forward from the first tail bound as follows:

$$\begin{aligned} \mathbb{P} \left[\left| \sum_i \beta_i x_{0i} \right| > \frac{\lambda}{10} \right] &\leq \mathbb{P} [|\beta_\tau^* \mathbf{x}_\tau| + |\beta_{\tau^c}^* \mathbf{x}_{\tau^c}| > \lambda/10] \\ &\leq \mathbb{P} [\mathbf{x}_\tau \neq \mathbf{0}] + \mathbb{P} [\mathbf{x}_\tau = \mathbf{0}] \cdot \mathbb{P} [|\beta_{\tau^c}^* \mathbf{x}_{\tau^c}| > \lambda/10] \\ &\leq \theta |\tau| + 2\theta. \end{aligned} \quad (\text{E.27})$$

Corollary E.7 ($|\langle \beta_{\tau \setminus (0)}, \mathbf{x}_{0, \tau \setminus (0)} \rangle|$ is small near shifts). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and $|\tau|, c_\mu$ such that $(\mathbf{a}_0, \theta, |\tau|)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$, then if $c_\mu \leq \frac{1}{10}$, for any \mathbf{a} such that $|\beta_{(1)}| \leq \frac{\lambda}{4 \log \theta^{-1}}$, we have*

$$\mathbb{P} \left[\left| \sum_{i \in \tau \setminus (0)} \beta_i x_{0i} \right| > \frac{2\lambda}{5} \right] \leq 2\theta \quad (\text{E.28})$$

Proof. For the last tail bound, write $\mathbf{x} = \omega \circ \mathbf{g}$. Wlog define β_0 be the largest correlation $\beta_{(0)}$, define random variables $s' = \langle \beta_{\tau \setminus \{0\}}, \mathbf{x}_{\tau \setminus \{0\}} \rangle$. Firstly most of the entries of \mathbf{x}_τ would be zero since via Bernstein inequality with $\theta |\tau| < 0.1$:

$$\begin{aligned} \mathbb{P} \left[\sum_{i \in \tau} \omega_i > \log \theta^{-1} \right] &\leq \mathbb{P} \left[\sum_{i \in \tau} \omega_i > \theta |\tau| + 0.9 \log \theta^{-1} \right] \\ &\leq \exp \left(\frac{-0.9^2 \log^2 \theta^{-1}}{2(\theta |\tau| + 0.9 \log \theta^{-1}/3)} \right) \leq \theta \end{aligned} \quad (\text{E.29})$$

thus with probability at least $1 - \theta$, we can write s' as a Gaussian r.v. with variation bounded as $\mathbb{E} s'^2 \leq \mathbb{E} \left[\sum_{i=1}^{\log \theta^{-1}} \beta_i \mathbf{g}_i \right]^2 = \log \theta^{-1} \beta_{(1)}^2$, then via Gaussian tail bound [Lemma N.1](#):

$$\begin{aligned} \mathbb{P} [s' > 0.4\lambda] &\leq \mathbb{P} \left[|g| > \frac{0.4\lambda}{\sqrt{\log \theta^{-1}} |\beta_{(1)}|} \right] + \mathbb{P} \left[\sum_{i \in \tau} \omega_i > \log \theta^{-1} \right] \\ &\leq \frac{2}{\sqrt{2\pi}} \exp(-1.2 \log \theta^{-1}) + \theta \leq 2\theta, \end{aligned} \quad (\text{E.30})$$

F Euclidean gradient as soft-thresholding in shift space

In this section, we will study the Euclidean gradient (B.6), by deriving bounds showing that the χ operator approximates a soft-thresholding function in shift space (Lemma F.2 and Corollary F.4). Furthermore, we will show the operator $\chi[\beta_i]$ is monotone in $|\beta_i|$ from Lemma F.3. A figure of visualized χ operator is shown in Figure 3.

To understand the χ operator, we shall first consider a simple case—when \mathbf{x}_0 is highly sparse. By definition of β from (B.3) we can see that β has a short support of size at most $2p - 1$, when \mathbf{x}_0 has support entries separated by at least $2p$, the entries of vector $\chi[\beta]_i$ become sum of independent random variables as:

$$\chi[\beta]_i = \left\langle s_{-i}[\mathbf{x}_0], \mathcal{S}_\lambda \left[\mathbf{x}_0 * \check{\beta} \right] \right\rangle \underset{\mathbf{x}_0 \text{ sep.}}{=} \left\langle s_{-i}[\mathbf{x}_0], \mathcal{S}_\lambda [\beta_i s_{-i}[\mathbf{x}_0]] \right\rangle = \sum_{j \in \text{supp}(\mathbf{x}_0)} \mathbf{g}_j \cdot \mathcal{S}_\lambda [\mathbf{g}_j \cdot \beta_i]$$

where $(\mathbf{g}_j)_{j \in [n]}$ are standard Gaussian r.v.s.

The following lemma describes the behavior of the summands in the above expression:

Lemma F.1 (Gaussian smoothed soft-thresholding). *Let $g \sim \mathcal{N}(0, 1)$. Then for every $b, s \in \mathbb{R}$ and $\lambda > 0$,*

$$\mathbb{E}_g \left[g \mathcal{S}_\lambda [b \cdot g + s] \right] = b(1 - \text{erf}_b(\lambda, s)), \quad (\text{F.1})$$

where

$$\text{erf}_b(\lambda, s) = \frac{1}{2} \text{erf} \left(\frac{\lambda + s}{\sqrt{2}|b|} \right) + \frac{1}{2} \text{erf} \left(\frac{\lambda - s}{\sqrt{2}|b|} \right). \quad (\text{F.2})$$

Furthermore, for $s = 0$, $b \in [-1, 1]$ and $\varepsilon \in (0, 1/4)$, letting $\sigma = \text{sign}(b)$ we have

$$\sigma \mathcal{S}_{\nu'_2 \lambda} [b] \leq \sigma \mathbb{E}_g \left[g \mathcal{S}_\lambda [b \cdot g] \right] \leq \sigma \mathcal{S}_{\nu'_1(\varepsilon) \lambda} [b] + \varepsilon \quad (\text{F.3})$$

where $\nu'_1(\varepsilon) = 1/(2\sqrt{-\log \varepsilon})$ and $\nu'_2 = \sqrt{2/\pi}$.

Proof. Wlog assume $b > 0$. Write f as the pdf of standard Gaussian distribution. With integral by parts:

$$\int_{-\infty}^t t' f(t') dt' = -f(t), \quad \int_{-\infty}^t t'^2 f(t') dt' = \frac{1}{2} \text{erf} \left(\frac{t}{\sqrt{2}} \right) - t f(t)$$

Integrating, we obtain

$$\mathbb{E} \left[g \mathcal{S}_\lambda [b \cdot g + s] \right] = \int_{t \geq \frac{\lambda - s}{b}} (bt^2 - (\lambda - s)t) f(t) dt + \int_{t \leq -\frac{\lambda + s}{b}} (bt^2 + (\lambda + s)t) f(t) dt,$$

by writing $L = \lambda - s$, the integral of first summand

$$\begin{aligned} \int_{t \geq \frac{L}{b}} (bt^2 - Lt) f(t) dt &= b \left[\frac{1}{2} - \frac{1}{2} \text{erf} \left(\frac{L}{\sqrt{2}b} \right) + \frac{L}{b} f \left(\frac{L}{b} \right) \right] - L f \left(\frac{L}{b} \right) \\ &= \frac{b}{2} - \frac{b}{2} \text{erf} \left(\frac{L}{\sqrt{2}b} \right), \end{aligned}$$

and similarly for the second summand, which gives

$$\mathbb{E} \left[g \mathcal{S}_\lambda [b \cdot g + s] \right] = \frac{b}{2} - \frac{b}{2} \text{erf} \left(\frac{\lambda - s}{\sqrt{2}b} \right) + \frac{b}{2} - \frac{b}{2} \text{erf} \left(\frac{\lambda + s}{\sqrt{2}b} \right) = b(1 - \text{erf}_b(\lambda, s))$$

For $b < 0$, alternatively we have

$$\mathbb{E}\left[gS_\lambda[-|b| \cdot g + s]\right] = -\mathbb{E}[gS_\lambda[|b| \cdot g - s]] = -|b|(1 - \text{erf}_b(\lambda, -s)) = b(1 - \text{erf}_b(\lambda, s)),$$

To show (F.3), via definition of error function, for $x > 0$, we know:

$$\min \left\{ 1 - \varepsilon, \frac{1 - \varepsilon}{\sqrt{\log(1/\varepsilon)}} x \right\} \leq \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \leq \frac{2x}{\sqrt{\pi}} \quad (\text{F.4})$$

where the lower bound is derived by first knowing erf is increasing thus for all $x > \sqrt{\log(1/\varepsilon)}$,

$$\text{erf}(x) \geq 1 - e^{-x^2} \geq 1 - e^{\log \varepsilon} = 1 - \varepsilon$$

and from concavity of erf we have for $0 < x < \sqrt{\log(1/\varepsilon)} = T$,

$$\text{erf}(x) \geq \frac{\text{erf}(T) - \text{erf}(0)}{T - 0} x + \text{erf}(0) \geq \frac{1 - \varepsilon}{\sqrt{\log(1/\varepsilon)}} x.$$

Lastly plug (F.4) into (F.1) and apply condition $|b| \leq 1$ and $\varepsilon < 1/4$ we have

$$\begin{aligned} |b| - \sqrt{\frac{2}{\pi}} \lambda &\leq |b| - |b| \text{erf} \left(\frac{\lambda}{\sqrt{2}|b|} \right) \\ &\leq \max \left\{ |b| \varepsilon, |b| - \frac{\lambda(1 - \varepsilon)}{\sqrt{2} \log(1/\varepsilon)} \right\} \leq \max \left\{ \varepsilon, |b| - \frac{\lambda}{2\sqrt{\log(1/\varepsilon)}} \right\}, \end{aligned}$$

which completes the proof. ■

This lemma establishes when \mathbf{x}_0 is separated, then χ is soft thresholding operator on β with threshold about $\lambda/2$. This phenomenon extends beyond the separated case, as long as when \mathbf{x}_0 is sufficiently sparse (when Definition E.1 holds). Recall that $\chi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined as

$$\chi[\beta] = \check{C}_{\mathbf{x}_0} S_\lambda \left[\check{C}_{\mathbf{x}_0} \beta \right]. \quad (\text{F.5})$$

The following lemma bounds its expectation:

Lemma F.2 (Expectation of $\chi(\beta)$). *Let $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ and $\lambda > 0$, then for every $\mathbf{a} \in \mathbb{S}^{p-1}$ and every $i \in [n]$, define the operator χ as in (F.5), then*

$$n^{-1} \mathbb{E} \chi[\beta]_i = \theta \beta_i (1 - \mathbb{E}_{\mathbf{s}_i} \text{erf}_{\beta_i}(\lambda, \mathbf{s}_i)) \quad (\text{F.6})$$

where $\mathbf{s}_i = \sum_{\ell \neq i} \beta_\ell \mathbf{x}_{0\ell}$. Suppose $(\mathbf{a}_0, \theta, |\tau|)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$ and $\lambda = c_\lambda / \sqrt{|\tau|}$ for some $c_\lambda > 1/5$ and $\sigma_i = \text{sign}(\beta_i)$, then there exists some numerical constant \bar{c} such that if $c_\mu \leq \bar{c}$ then for every $\mathbf{a} \in \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and every $i \in [n]$, (F.6) has upper bound

$$\sigma_i n^{-1} \mathbb{E} \chi[\beta]_i \leq \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta]_i} := \begin{cases} 4\theta^2 |\tau| |\beta_i| & |\beta_i| < \nu_1 \lambda \\ \theta (|\beta_i| - \nu_1 \lambda / 2) & |\beta_i| \geq \nu_1 \lambda \end{cases}, \quad (\text{F.7})$$

and lower bound

$$\sigma_i n^{-1} \mathbb{E} \chi[\beta]_i \geq \sigma_i n^{-1} \underline{\mathbb{E} \chi[\beta]_i} =: \theta \mathcal{S}_{\nu_2 \lambda} [|\beta_i|], \quad (\text{F.8})$$

where $\nu_1 = 1 / (2\sqrt{\log \theta^{-1}})$, $\nu_2 = \sqrt{2/\pi}$.

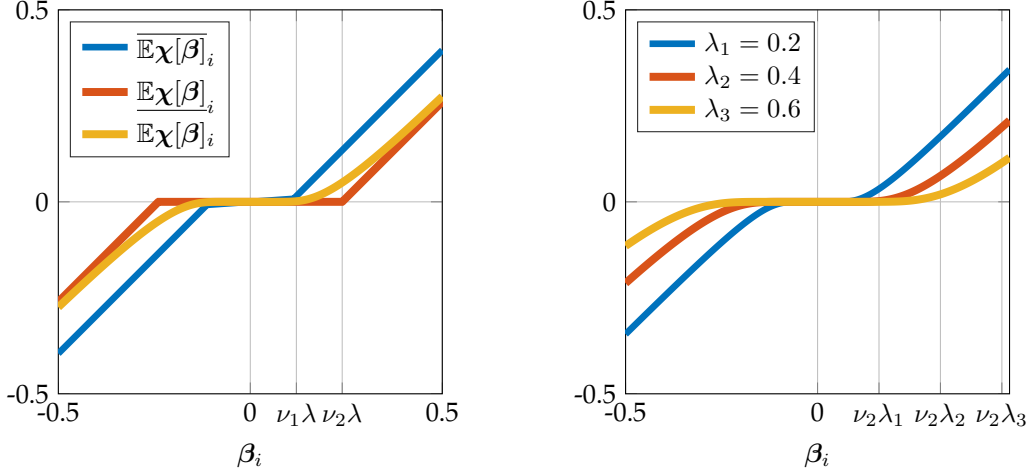


Figure 3: A numerical example of $\mathbb{E}\chi[\beta]_i$. We provide figures for the expectation of χ when entries of \mathbf{x}_0 are $2p$ -separated. Left: the yellow line is the function $\beta_i \rightarrow \beta_i (1 - \text{erf}_{\beta_i}(\lambda, 0))$ derived from (F.1), and the blue/red lines are its upper/lower bound (F.3) utilized in the analysis respectively. Right: functions of $\beta_i \rightarrow \beta_i (1 - \text{erf}_{\beta_i}(\lambda, 0))$ with different λ , the section of function of $\beta_i > \nu_2 \lambda$ are close to linear.

This lemma shows the expectation of $\chi[\beta]_i$ acts like a shrinkage operation on $|\beta_i|$: for large $|\beta_i|$, it acts like a soft thresholding operation, and for small $|\beta_i|$, it reduces $|\beta_i|$ by multiplying a very small number $4\theta|\tau| \ll 1$. We rigorously prove this segmentation of χ operator as follows:

Proof. First, since $s_i[\mathbf{x}_0] \equiv_d s_j[\mathbf{x}_0]$,

$$\begin{aligned} \chi[\beta]_i &= \mathbf{e}_i^* \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda [\check{\mathbf{C}}_{\mathbf{x}_0} \boldsymbol{\beta}] = \langle s_{-i}[\mathbf{x}_0], \mathcal{S}_\lambda [\mathbf{x}_0 * \check{\boldsymbol{\beta}}] \rangle \\ &\equiv_d \langle s_{-j}[\mathbf{x}_0], \mathcal{S}_\lambda [s_{i-j}[\mathbf{x}_0] * \check{\boldsymbol{\beta}}] \rangle = \chi[s_{j-i}[\boldsymbol{\beta}]]_j \end{aligned}$$

Thus wlog let us consider $i = 0$ and write \mathbf{x} as \mathbf{x}_0 . The random variable $\chi[\beta]_0$ can be written sum of random variables as:

$$\chi[\beta]_0 = \left\langle \mathbf{x}, \mathcal{S}_\lambda \left[\beta_0 \mathbf{x}_0 + \sum_{\ell \neq 0} \beta_\ell s_{-\ell}[\mathbf{x}] \right] \right\rangle = \sum_{j \in [n]} \mathbf{x}_j \mathcal{S}_\lambda \left[\beta_0 \mathbf{x}_j + \sum_{\ell \neq 0} \beta_\ell \mathbf{x}_{j+\ell} \right],$$

and a random variable $Z_j(\boldsymbol{\beta})$ is defined as

$$Z_j(\boldsymbol{\beta}) = \mathbf{x}_j \mathcal{S}_\lambda \left[\beta_0 \mathbf{x}_j + \sum_{\ell \in [\pm p] \setminus 0} \beta_\ell \mathbf{x}_{j+\ell} \right], \quad (\text{F.9})$$

gives $\chi[\beta]_0 = \sum_{j \in [n]} Z_j(\boldsymbol{\beta})$ as sum of r.v.s. of same distribution and thus $n^{-1} \mathbb{E}\chi[\beta]_0 = \mathbb{E}Z_0(\boldsymbol{\beta})$. Define a random variable $\mathbf{s}_0 = \sum_{\ell \neq 0} \beta_\ell \mathbf{x}_\ell$, which is independent of \mathbf{x}_0 . From Lemma F.1, we can conclude

$$n^{-1} \mathbb{E}\chi[\beta]_0 = \mathbb{E}_{\mathbf{x}_0, \mathbf{s}_0} \mathbf{x}_0 \mathcal{S}_\lambda [\beta_0 \mathbf{x}_0 + \mathbf{s}_0] = \theta \beta_0 (1 - \mathbb{E}_{\mathbf{s}_0} \text{erf}_{\beta_0}(\lambda, \mathbf{s}_0)) \quad (\text{F.10})$$

so that (F.6) holds for $i = 0$, and hence for all i .

1. Upper bound of $\mathbb{E}Z$ Wlog assume $\beta_0 \geq 0$ and write $Z = Z_0$. We derive the upper bound on $\mathbb{E}Z$ in two pieces.

(1). First, since $\mathbb{E}\mathbf{x}_0\mathcal{S}_\lambda [0 \cdot \mathbf{x}_0 + \mathbf{s}_0] = 0$, we have

$$\begin{aligned}
\mathbb{E}Z(\boldsymbol{\beta}) &\leq \boldsymbol{\beta}_0 \sup_{\beta \in [0, \boldsymbol{\beta}_0]} \frac{d}{d\beta} \mathbb{E}_{\mathbf{x}_0, \mathbf{s}_0} \mathbf{x}_0 \mathcal{S}_\lambda [\beta \mathbf{x}_0 + \mathbf{s}_0] \\
&= \theta \boldsymbol{\beta}_0 \sup_{\beta \in [0, \boldsymbol{\beta}_0]} \frac{d}{d\beta} \int_{|\beta g + \mathbf{s}_0| > \lambda} g(\beta g + \mathbf{s}_0 - \text{sign}(\beta g + \mathbf{s}_0) \cdot \lambda) d\mu(g) d\mu(\mathbf{s}_0) \\
&= \theta \boldsymbol{\beta}_0 \sup_{\beta \in [0, \boldsymbol{\beta}_0]} \mathbb{E}_{g, \mathbf{s}_0} [g^2 \mathbf{1}_{\{|\beta g + \mathbf{s}_0| > \lambda\}}] \\
&\leq \theta \boldsymbol{\beta}_0 \sup_{\beta \in [0, \boldsymbol{\beta}_0]} \mathbb{E}_{g, \mathbf{s}_0} \left[g^2 \left(\mathbf{1}_{\{|\beta g| > \frac{9\lambda}{10}\}} + \mathbf{1}_{\{|\mathbf{s}_0| > \frac{\lambda}{10}\}} \right) \right] \\
&\leq \theta \boldsymbol{\beta}_0 \left((\mathbb{E}g^6)^{1/3} \mathbb{P}[|\boldsymbol{\beta}_0 g| > (9\lambda/10)]^{2/3} + \mathbb{P}[|\mathbf{s}_0| > \lambda/10] \right)
\end{aligned} \tag{F.11}$$

We bound the tail probability of \mathbf{s}_0 using [Corollary E.6](#) where

$$\mathbb{P}[|\mathbf{s}_0| > \lambda/10] \leq \mathbb{P}[|\sum_i \boldsymbol{\beta}_i \mathbf{x}_i| > \lambda/10] \leq \theta |\boldsymbol{\tau}| + 2\theta \leq 3\theta |\boldsymbol{\tau}|. \tag{F.12}$$

On the other hand, the first term in [\(F.11\)](#) can be derived by pdf of Gaussian r.v. [Lemma N.1](#) as:

$$\begin{aligned}
(\mathbb{E}g^6)^{1/3} \mathbb{P}[|\boldsymbol{\beta}_0 g| > (9\lambda/10)]^{2/3} &\leq \sqrt[3]{15} \left(\frac{10\boldsymbol{\beta}_0}{9\lambda\sqrt{2\pi}} \right)^{2/3} \exp\left(-\frac{\lambda^2}{4\boldsymbol{\beta}_0^2}\right) \\
&\leq \frac{3}{2} \left(\frac{\boldsymbol{\beta}_0}{\lambda} \right)^{2/3} \exp\left(-\frac{\lambda^2}{4\boldsymbol{\beta}_0^2}\right).
\end{aligned} \tag{F.13}$$

Combine [\(E.26\)](#), [\(F.13\)](#), when $\boldsymbol{\beta}_0 < \nu_1 \lambda$, we know $e^{-\frac{\lambda^2}{4\boldsymbol{\beta}_0^2}} \leq e^{\log \theta} \leq \theta |\boldsymbol{\tau}|$. The first type of upper bound $\mathbb{E}Z$ is derived as

$$\forall \boldsymbol{\beta}_0 \in [0, \nu_1 \lambda], \quad \mathbb{E}Z(\boldsymbol{\beta}) \leq \theta \boldsymbol{\beta}_0 \left(\frac{3}{2} \nu_1^{2/3} \exp\left(-\frac{\lambda^2}{4\boldsymbol{\beta}_0^2}\right) + 3\theta |\boldsymbol{\tau}| \right) \leq 4\theta^2 |\boldsymbol{\tau}| \boldsymbol{\beta}_0. \tag{F.14}$$

(2). The second type of upper bound can be derived directly from [Lemma F.1](#):

$$\begin{aligned}
\mathbb{E}Z(\boldsymbol{\beta}) &\leq \mathbb{E}_{\mathbf{x}_0} \mathbb{E}_{\mathbf{s}_0} \mathbf{x}_0 \mathcal{S}_\lambda [\boldsymbol{\beta}_0 \mathbf{x}_0 + \mathbf{s}_0] \leq \mathbb{E}_{\mathbf{x}_0} \mathbf{x}_0 \mathcal{S}_\lambda [\boldsymbol{\beta}_0 \mathbf{x}_0] + \mathbb{E}_{\mathbf{x}_0} |\mathbf{x}_0| \mathbb{E}_{\mathbf{s}_0} |\mathbf{s}_0| \\
&\leq \theta \cdot \left(\mathcal{S}_{\nu_1 \lambda} [\boldsymbol{\beta}_0] + \varepsilon + \sqrt{2/\pi} \cdot \mathbb{E}|\mathbf{s}_0| \right),
\end{aligned} \tag{F.15}$$

where $\mathbb{E}|\mathbf{s}_0|$ can be bounded with $\|\boldsymbol{\beta}\|_2$ and $\theta |\boldsymbol{\tau}| < c_\mu \theta_{\log}$ from [Lemma E.5](#). When $c_\mu < \frac{1}{10}$, observe that

$$\mathbb{E}|\mathbf{s}_0| \leq \sqrt{\sum_\ell \mathbb{E}\mathbf{x}_\ell^2 \boldsymbol{\beta}_\ell^2} \leq \sqrt{\theta} (\|\boldsymbol{\beta}_\boldsymbol{\tau}\|_2 + \|\boldsymbol{\beta}_{\boldsymbol{\tau}^c}\|_2) \leq \sqrt{\theta} (1 + c_\mu) + \frac{c_\mu \theta_{\log}}{|\boldsymbol{\tau}|} \leq \frac{2c_\mu \theta_{\log}}{\sqrt{|\boldsymbol{\tau}|}}. \tag{F.16}$$

Now choose $\varepsilon = \theta \leq \frac{c_\mu \theta_{\log}}{|\boldsymbol{\tau}|}$, so that $\nu'_1 = \nu_1 = \frac{\sqrt{\theta_{\log}}}{2}$ in [\(F.15\)](#). Since $c_\mu < \frac{c_\lambda}{25}$ we gain

$$\begin{aligned}
\mathbb{E}Z(\boldsymbol{\beta}) &\leq \theta \left(\mathcal{S}_{\nu_1 \lambda} [\boldsymbol{\beta}_0] + \frac{c_\mu \theta_{\log}}{|\boldsymbol{\tau}|} + \sqrt{\frac{2}{\pi}} \cdot \frac{2c_\mu \theta_{\log}}{\sqrt{|\boldsymbol{\tau}|}} \right) \leq \theta \left(\mathcal{S}_{\nu_1 \lambda} [\boldsymbol{\beta}_0] + \frac{3c_\mu \theta_{\log}}{\sqrt{|\boldsymbol{\tau}|}} \right) \\
&\leq \theta \left(\mathcal{S}_{\nu_1 \lambda} [\boldsymbol{\beta}_0] + \frac{\sqrt{\theta_{\log}}}{5} \lambda \right) \leq \theta \left(\mathcal{S}_{\nu_1 \lambda} [\boldsymbol{\beta}_0] + \frac{1}{2} \nu_1 \lambda \right)
\end{aligned} \tag{F.17}$$

(3). Combine both [\(F.14\)](#) and [\(F.17\)](#), we can thus conclude that

$$\mathbb{E}Z(\boldsymbol{\beta}) := \overline{\mathbb{E}Z(\boldsymbol{\beta})} \leq \begin{cases} 4\theta^2 |\boldsymbol{\tau}| \boldsymbol{\beta}_0 & \boldsymbol{\beta}_0 \leq \nu_1 \lambda \\ \theta (\boldsymbol{\beta}_0 - \frac{\nu_1}{2} \lambda) & \boldsymbol{\beta}_0 > \nu_1 \lambda \end{cases}. \tag{F.18}$$

2. (Lower bound of $\mathbb{E}Z$) On the other hand, for the lower bound for $\mathbb{E}Z$, use the fact that $\text{erf}_\beta(\lambda, s)$ is concave in s_0 , we have

$$\begin{aligned}\mathbb{E}Z(\boldsymbol{\beta}) &= \mathbb{E}_{\mathbf{s}_0} \mathbb{E}_{\mathbf{x}_0} \mathbf{x}_0 \mathcal{S}_\lambda [\boldsymbol{\beta}_0 \mathbf{x}_0 + \mathbf{s}_0] \\ &= \theta \cdot \mathbb{E}_{\mathbf{s}_0} \left[\boldsymbol{\beta}_0 - \frac{\boldsymbol{\beta}_0}{2} \cdot \text{erf} \left(\frac{\lambda - \mathbf{s}_0}{\sqrt{2} |\boldsymbol{\beta}_0|} \right) - \frac{\boldsymbol{\beta}_0}{2} \cdot \text{erf} \left(\frac{\lambda + \mathbf{s}_0}{\sqrt{2} |\boldsymbol{\beta}_0|} \right) \right] \\ &\geq \theta \left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0 \cdot \text{erf} \left(\frac{\lambda}{\sqrt{2} |\boldsymbol{\beta}_0|} \right) \right) \geq \theta \cdot \mathcal{S}_{\nu_2 \lambda} [\boldsymbol{\beta}_0] =: \underline{\mathbb{E}Z(\boldsymbol{\beta})}.\end{aligned}\tag{F.19}$$

The proof of $\boldsymbol{\beta}_0 < 0$ is in the same vein. For cases of $i \neq 0$, since $\chi[\boldsymbol{\beta}]_i \equiv_d \chi[s_{-i}[\boldsymbol{\beta}]]_0$, replace $\boldsymbol{\beta}_0$ with $\boldsymbol{\beta}_i$ we obtain the desired result. \blacksquare

Another convenient fact of $\mathbb{E}\chi[\boldsymbol{\beta}]_i$ is that it is monotone increasing w.r.t. $|\boldsymbol{\beta}_i|$. The monotonicity is clear in [Figure 3](#); it is demonstrated rigorously with the following lemma:

Lemma F.3 (Monotonicity of $\mathbb{E}\chi(\boldsymbol{\beta})$). *Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and $|\boldsymbol{\tau}|, c_\mu$ such that $(\mathbf{a}_0, \theta, |\boldsymbol{\tau}|)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda / \sqrt{|\boldsymbol{\tau}|}$ in φ_{ℓ^1} where $c_\lambda \in [0, \frac{1}{4}]$, then there exists some numerical constant $\bar{c} > 0$, such that if $c_\mu < \bar{c}$, the expectation $|\mathbb{E}\chi[\boldsymbol{\beta}]_i|$ is monotone increasing in $|\boldsymbol{\beta}_i|$. In other words, if $|\boldsymbol{\beta}_i| > |\boldsymbol{\beta}_j|$ then*

$$\sigma_i \mathbb{E}\chi[\boldsymbol{\beta}]_i \geq \sigma_j \mathbb{E}\chi[\boldsymbol{\beta}]_j \tag{F.20}$$

where $\sigma_i = \text{sign}(\boldsymbol{\beta}_i)$.

The proof first operate simple calculus and then followed by studying cases of $|\boldsymbol{\beta}_i| - |\boldsymbol{\beta}_j|$ when either it is smaller are larger then λ .

Proof. 1. (Monotonicity by gradient negativity) Wlog assume $\boldsymbol{\beta}_i > \boldsymbol{\beta}_j > 0$, and from [Lemma F.2](#) we can write $(n\theta)^{-1} \mathbb{E}\chi[\boldsymbol{\beta}]_i = \boldsymbol{\beta}_i (1 - \mathbb{E}_{\mathbf{s}_i} \text{erf}_{\boldsymbol{\beta}_i}(\lambda, \mathbf{s}_i))$. Consider $t \in [0, 1]$ and define $\ell(t) = t\boldsymbol{\beta}_i - t\boldsymbol{\beta}_j$. Write the random variable $\mathbf{s}_{ij} = \sum_{\ell \neq i, j} \boldsymbol{\beta}_\ell \mathbf{x}_\ell$. Define h as a function of t such that

$$\begin{aligned}h(t) &= \mathbb{E}_{\mathbf{x}, \mathbf{s}_{ij}} \left[((1-t)\boldsymbol{\beta}_i + t\boldsymbol{\beta}_j) (1 - \text{erf}_{(1-t)\boldsymbol{\beta}_i + t\boldsymbol{\beta}_j}(\lambda, ((1-t)\boldsymbol{\beta}_j + t\boldsymbol{\beta}_i)x + \mathbf{s}_{ij})) \right] \\ &= \mathbb{E}_{\mathbf{x}, \mathbf{s}_{ij}} \left[(\boldsymbol{\beta}_i - \ell(t)) (1 - \text{erf}_{\boldsymbol{\beta}_i - \ell(t)}(\lambda, x \cdot (\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij})) \right].\end{aligned}\tag{F.21}$$

Notice that $\mathbb{E}\chi[\boldsymbol{\beta}]_i = h(0)$ and $\mathbb{E}\chi[\boldsymbol{\beta}]_j = h(1)$ respectively, thus it suffices to prove $h'(t) < 0$ for all $t \in [0, 1]$. Write f as pdf of standard Gaussian r.v. where

$$\text{erf}_\beta(\lambda, \mathbf{s}_{ij}) = \int_0^{\frac{\lambda + \mathbf{s}_{ij}}{\beta}} f(z) dz + \int_0^{\frac{\lambda - \mathbf{s}_{ij}}{\beta}} f(z) dz,$$

and use chain rule:

$$\begin{aligned}h'(t) &= \mathbb{E}_{\mathbf{x}, \mathbf{s}_{ij}} \left[(\boldsymbol{\beta}_j - \boldsymbol{\beta}_i) (1 - \text{erf}_{\boldsymbol{\beta}_i - \ell(t)}(\lambda, x \cdot (\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij})) \right. \\ &\quad - (\boldsymbol{\beta}_i - \ell(t)) \cdot \frac{d}{dt} \left(\frac{\lambda + x \cdot (\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} \right) \cdot f \left(\frac{\lambda + x \cdot (\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} \right) \\ &\quad \left. - (\boldsymbol{\beta}_i - \ell(t)) \cdot \frac{d}{dt} \left(\frac{\lambda - x \cdot (\boldsymbol{\beta}_j + \ell(t)) - \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} \right) \cdot f \left(\frac{\lambda - x \cdot (\boldsymbol{\beta}_j + \ell(t)) - \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} \right) \right] \\ &= (\boldsymbol{\beta}_j - \boldsymbol{\beta}_i) \mathbb{E}_{\mathbf{x}, \mathbf{s}_{ij}} \left[1 - \text{erf}_{\boldsymbol{\beta}_i - \ell(t)}(\lambda, x \cdot (\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij}) \right. \\ &\quad \left. + \underbrace{\left(\frac{\lambda + x(\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} + x \right)}_{z_{\lambda_+}} \cdot f \left(\frac{\lambda + x(\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij}}{\boldsymbol{\beta}_i - \ell(t)} \right) \right]\end{aligned}$$

$$\begin{aligned}
& + \left(\frac{\lambda - x(\beta_j + \ell(t)) - \mathbf{s}_{ij}}{\beta_i - \ell(t)} - x \right) \cdot \underbrace{f \left(\frac{\lambda - x(\beta_j + \ell(t)) - \mathbf{s}_{ij}}{\beta_i - \ell(t)} \right)}_{z_{\lambda_-}} \\
& = (\beta_j - \beta_i) \mathbb{E}_{x, \mathbf{s}_{ij}} \left[1 - \int_0^{z_{\lambda_+}} f(z) dz - \int_0^{z_{\lambda_-}} f(z) dz + (z_{\lambda_+} + x)f(z_{\lambda_+}) + (z_{\lambda_-} - x)f(z_{\lambda_-}) \right]. \quad (\text{F.22})
\end{aligned}$$

Consider the term only related to z_{λ_+} , condition on cases that it is either positive or negative, observe that

$$\begin{cases} \mu_{+-} := \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_+} \leq 0} \left[\int_0^{z_{\lambda_+}} f(z) dz - z_{\lambda_+} f(z_{\lambda_+}) \right] = \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_+} \leq 0} \left[- \int_0^{-z_{\lambda_+}} f(z) dz - z_{\lambda_+} f(z_{\lambda_+}) \right] \leq 0 \\ \mu_{++} := \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_+} > 0} \left[\int_0^{z_{\lambda_+}} f(z) dz - z_{\lambda_+} f(z_{\lambda_+}) \right] \leq \min \left\{ \frac{1}{2}, \frac{1}{\sqrt{2\pi}} \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_+} > 0} z_{\lambda_+} \right\} \end{cases},$$

where the negativity of the first equation can be observed by writing $v = -z_{\lambda_+}$ and take derivative:

$$\begin{cases} - \int_0^v f(z) dz + v \cdot f(v) = 0 & v = 0 \\ \frac{d}{dv} \left\{ - \int_0^v f(z) dz + v \cdot f(v) \right\} = -f(v) + f(v) + v \cdot f'(v) < 0 & v > 0 \end{cases};$$

and similarly for z_{λ_-} :

$$\begin{cases} \mu_{--} := \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_-} \leq 0} \left[\int_0^{z_{\lambda_-}} f(z) dz - z_{\lambda_-} f(z_{\lambda_-}) \right] \leq 0 \\ \mu_{-+} := \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_-} > 0} \left[\int_0^{z_{\lambda_-}} f(z) dz - z_{\lambda_-} f(z_{\lambda_-}) \right] \leq \min \left\{ \frac{1}{2}, \frac{1}{\sqrt{2\pi}} \mathbb{E}_{x, \mathbf{s}_{ij} | z_{\lambda_-} > 0} z_{\lambda_-} \right\} \end{cases},$$

then combine every term to (F.22) using tower property and from assumption $\beta_j - \beta_i < 0$ we obtain

$$\begin{aligned}
(\text{F.22}) & \leq (\beta_j - \beta_i) \left(1 - \mathbb{P}[z_{\lambda_+} > 0] \cdot \mu_{++} \right. \\
& \quad \left. - \mathbb{P}[z_{\lambda_-} > 0] \cdot \mu_{-+} + \mathbb{E}_{x, \mathbf{s}_{ij}} [x(f(z_{\lambda_+}) - f(z_{\lambda_-}))] \right) \\
& \leq (\beta_j - \beta_i) \left(1 - \min \left\{ \frac{\mathbb{P}[z_{\lambda_+} > 0]}{2}, \frac{\mathbb{E}|z_{\lambda_+}|}{\sqrt{2\pi}} \right\} \right. \\
& \quad \left. - \min \left\{ \frac{\mathbb{P}[z_{\lambda_-} > 0]}{2}, \frac{\mathbb{E}|z_{\lambda_-}|}{\sqrt{2\pi}} \right\} - \frac{\theta}{\sqrt{2\pi}} \cdot \mathbb{E}|g| \right), \quad (\text{F.23})
\end{aligned}$$

where g is standard Gaussian r.v..

2. (Cases of varying β_i, β_j) Let $c_\lambda < \frac{1}{4}$. Suppose $\beta_i - \ell(t) \leq \frac{1}{4\sqrt{|\tau|}}$. Recall that $\|\beta_\tau\|_2^2 \geq 1 - 3c_\mu$. We are going to show there is at least one of the entry $\beta_* \in \{\beta_r\}_{r \in \tau \neq i, j} \uplus \{\beta_j + \ell(t)\}$ is greater than $\frac{0.85}{\sqrt{|\tau|}}$. First, if both $i, j \notin \tau$, the lower bound is immediate since $\beta_*^2 = \|\beta_\tau\|_\infty^2 > \frac{1-3c_\mu}{|\tau|}$. On the other hand if at least one of i, j is in τ and all other β_τ entries are small where $\|\beta_{\tau \setminus \{i, j\}}\|_\infty^2 < \frac{1-3c_\mu}{|\tau|}$, then we know via norm inequalities,

$$(\beta_i + \beta_j)^2 > \beta_i^2 + \beta_j^2 > \|\beta_\tau\|_2^2 - (|\tau| - 1) \|\beta_{\tau \setminus \{i, j\}}\|_\infty^2 \geq \frac{1 - 3c_\mu}{|\tau|}, \quad (\text{F.24})$$

which implies if $c_\mu < \frac{1}{100}$,

$$\beta_* = \beta_j + \ell(t) = (\beta_i + \beta_j) - (\beta_i - \ell(t)) \geq \frac{\sqrt{1 - 3c_\mu}}{\sqrt{|\tau|}} - \frac{1}{4\sqrt{|\tau|}} \geq \frac{0.72}{\sqrt{|\tau|}}. \quad (\text{F.25})$$

In this case, adopt result from Corollary E.6 such that $\mathbb{P}[\|\sum \beta_\ell x_\ell\| > \lambda/10] \leq 3\theta|\tau| \leq .01$, we have

$$\mathbb{P}[z_{\lambda_-} > 0] = \mathbb{P}[z_{\lambda_+} > 0] = 1 - \mathbb{P}[x(\beta_j + \ell(t)) + \mathbf{s}_{ij} < -\lambda]$$

$$\begin{aligned}
&\leq 1 - \mathbb{P}[\mathbf{x}_* \boldsymbol{\beta}_* < -11\lambda/10] \cdot \mathbb{P}[x(\boldsymbol{\beta}_j + \ell(t)) + \mathbf{s}_{ij} - \mathbf{x}_* \boldsymbol{\beta}_* < \lambda/10] \\
&\leq 1 - \theta \cdot \mathbb{P}\left[\mathbf{g}_* \cdot \frac{0.72}{\sqrt{|\boldsymbol{\tau}|}} < \frac{-11c_\lambda}{10\sqrt{|\boldsymbol{\tau}|}}\right] \cdot \left(1 - \mathbb{P}\left[\sum \boldsymbol{\beta}_\ell \mathbf{x}_\ell > \frac{\lambda}{10}\right]\right) \\
&\leq 1 - \theta \cdot \mathbb{P}[0.72 \cdot \mathbf{g}_* \leq -1.1 \cdot 0.25] \cdot (1 - 3c_\mu) \\
&\leq 1 - 0.35\theta.
\end{aligned} \tag{F.26}$$

On the other hand, when $\beta_i - \ell(t) \geq \frac{1}{4\sqrt{|\boldsymbol{\tau}|}}$, both $z_{\lambda_+}, z_{\lambda_-}$ are upper bounded via $|\boldsymbol{\tau}| \theta \leq \frac{1}{800}$ such as:

$$\begin{aligned}
\mathbb{E}_{x, \mathbf{s}_{ij}} |z_{\lambda_-}| = \mathbb{E}_{x, \mathbf{s}_{ij}} |z_{\lambda_+}| &\leq \mathbb{E}_{x, \mathbf{s}_{ij}} \frac{\lambda + |x(\boldsymbol{\beta}_j + \ell(t)) - \mathbf{s}_{ij}|}{\beta_i - \ell(t)} \\
&\leq 1 + 4\sqrt{|\boldsymbol{\tau}|} \cdot \left(\mathbb{E}_{x, \mathbf{s}_{ij}} |x(\boldsymbol{\beta}_j + \ell(t)) - \mathbf{s}_{ij}|^2\right)^{1/2} \\
&\leq 1 + 4\sqrt{|\boldsymbol{\tau}| \theta} \|\boldsymbol{\beta}\|_2 \leq 1 + 4\sqrt{|\boldsymbol{\tau}| \theta} \left(1 + c_\mu + \frac{c_\mu}{\sqrt{\theta} |\boldsymbol{\tau}|}\right) \leq 1.2.
\end{aligned} \tag{F.27}$$

Combine (F.23), (F.26) we have

$$h'(t) \leq (\beta_j - \beta_i) \left(1 - 2 \cdot \frac{(1 - 0.35\theta)}{2} - \frac{\theta}{\sqrt{2\pi}} \cdot \sqrt{\frac{2}{\pi}}\right) \leq 0.03\theta(\beta_j - \beta_i) < 0, \tag{F.28}$$

and combine (F.23), (F.27) and $\theta < c_\mu$ we have

$$h'(t) \leq (\beta_j - \beta_i) \left(1 - 2 \cdot \frac{1.2}{\sqrt{2\pi}} - \frac{\theta}{\sqrt{2\pi}} \cdot \sqrt{\frac{2}{\pi}}\right) \leq 0.03(\beta_j - \beta_i) < 0, \tag{F.29}$$

which proves the monotonicity. \blacksquare

When the signal length of \mathbf{y} is sufficiently large, operator $\boldsymbol{\chi}$ will be enough close to its expected value.

Corollary F.4 (Finite sample deviation of $\boldsymbol{\chi}(\boldsymbol{\beta})$). *Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_{ℓ^1} for some $c_\lambda > 1/5$, then there exists some numerical constants $C, c, \bar{c} > 0$, such that if $n \geq Cp^5\theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - 3/n$, for every $\mathbf{a} \in \cup_{|\boldsymbol{\tau}| \leq k} \mathfrak{R}(\mathcal{S}_\boldsymbol{\tau}, \gamma(c_\mu))$ and every $i \in [n]$, we have:*

$$|n^{-1} \boldsymbol{\chi}[\boldsymbol{\beta}]_i - n^{-1} \mathbb{E} \boldsymbol{\chi}[\boldsymbol{\beta}]_i| \leq c\theta/p^{3/2}, \tag{F.30}$$

Proof. See [Appendix L.1](#) \blacksquare

G Euclidean Hessian as logic function in shift space

We can express the (pseudo) curvature (B.10) in direction $\mathbf{v} \in \mathbb{S}^{p-1}$ in terms of the correlation $\gamma = \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\nu}$ between \mathbf{v} and \mathbf{a}_0 , giving

$$\mathbf{v}^* \tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v} = -\gamma^* \check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{P}_I \check{\mathbf{C}}_{\mathbf{x}_0} \gamma,$$

where

$$I(\mathbf{a}) = \text{supp} \left(\mathcal{S}_\lambda \left[\check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\nu} \mathbf{a} \right] \right) = \left\{ i \in [n] \mid \left| \mathbf{x}_0 * \check{\boldsymbol{\beta}} \right|_i > \lambda \right\}. \tag{G.1}$$

The i -th diagonal entry of $\check{C}_{\mathbf{x}_0} \mathbf{P}_{I(\mathbf{a})} \check{C}_{\mathbf{x}_0}$ is

$$-\mathbf{e}_i^* \check{C}_{\mathbf{x}_0} \mathbf{P}_{I(\mathbf{a})} \check{C}_{\mathbf{x}_0} \mathbf{e}_i = -\left\| \mathbf{P}_{I(\mathbf{a})} \check{C}_{\mathbf{x}_0} \mathbf{e}_i \right\|_2^2 = -\left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0] \right\|_2^2, \quad (\text{G.2})$$

which is the core component for us to study the curvature of objective φ_{ℓ^1} . We illustrate the expectation of diagonal term of Hessian in [Lemma G.2](#) and [Corollary G.3](#), whose figure of visualized $\left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0] \right\|_2$ is shown in [Figure 3](#). Lastly, we also prove the off-diagonal terms $\mathbf{e}_i^* \check{C}_{\mathbf{x}_0} \mathbf{P}_{I(\mathbf{a})} \check{C}_{\mathbf{x}_0} \mathbf{e}_j$ of Hessian is likely inconsequential in calculation of curvature in [Lemma G.4](#).

We expect the Hessian to have stronger negative component in the $s_i[\mathbf{a}_0]$ direction as $\left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0] \right\|_2$ becomes larger. This term can be tremendously simplified when \mathbf{x}_0 is very sparse: suppose all entries of its support I_0 are separated by at least $2p - 1$ samples, then by implementing the definition of support from [\(G.1\)](#), we can derive

$$-\left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0] \right\|_2^2 = -\sum_{j \in I_0} \mathbf{x}_{0j}^2 \mathbf{1}_{\{|\sum_{\ell} \beta_{\ell} \mathbf{x}_{0(\ell+j-i)}| > \lambda\}} \underbrace{\quad}_{\text{sep.}} - \sum_{j \in I_0} \mathbf{g}_j^2 \mathbf{1}_{\{|\beta_j \mathbf{g}_j| > \lambda\}}, \quad (\text{G.3})$$

where $\mathbf{1}$ is the indicator function and \mathbf{g}_j are independent standard Gaussian r.v.s.. In expectation, the summands in [\(G.3\)](#) acts like a smoothed logic function on entry β_j :

Lemma G.1 (Gaussian smoothed indicator). *Let $g \sim \mathcal{N}(0, 1)$, then for any $b, s \in \mathbb{R}$ and $\lambda > 0$.*

$$\mathbb{E}_g \left[g^2 \mathbf{1}_{\{|b \cdot g + s| > \lambda\}} \right] = 1 - \text{erf}_b(\lambda, s) + f_b(\lambda, s), \quad (\text{G.4})$$

where

$$f_b(\lambda, s) = \frac{1}{\sqrt{2\pi}} \left[\left(\frac{\lambda + s}{|b|} \right) e^{-\frac{(\lambda+s)^2}{2b^2}} + \left(\frac{\lambda - s}{|b|} \right) e^{-\frac{(\lambda-s)^2}{2b^2}} \right]. \quad (\text{G.5})$$

Proof. The proof can be derived via same calculation of integrals in [Lemma F.1](#). ■

Although the definition [\(G.4\)](#) seems incomprehensible at first glance, we can actually interpret it as a smoothed indicator function which compares $|b|$ to the threshold $\sqrt{2/\pi}\lambda$. Once we assign $s = 0$, then we can see that $\mathbb{E} g^2 \mathbf{1}_{\{|b \cdot g| > \lambda\}}$ is be an increasing function of $|b|$. Moreover by assigning different values for $|b|$ we obtain:

$$\mathbb{E} g^2 \mathbf{1}_{\{|b \cdot g| > \lambda\}} \approx \begin{cases} 1, & |b| \approx 1 \\ 1/2, & |b| \approx \sqrt{2/\pi}\lambda \\ 0, & |b| \approx 0 \end{cases}. \quad (\text{G.6})$$

Relate [\(G.6\)](#) to [\(G.3\)](#), when $|\beta_i|$ is close to 1 then we expect $-\frac{1}{n\theta} \left\| \mathbf{P}_{I} s_{-i}[\mathbf{x}_0] \right\|_2^2$ to be close to -1 , and it increases to 0 as $|\beta_i|$ decreases, suggests that the Euclidean Hessian at point \mathbf{a} has stronger negative component at $s_i[\mathbf{a}_0]$ direction if $|\langle \mathbf{a}, s_i[\mathbf{a}_0] \rangle|$ is larger. See [Figure 4](#) for a numerical example. This phenomenon can be extend beyond the idealistic separating case as follows:

Lemma G.2 (Expected Hessian diagonals). *Let $\mathbf{x}_0 \sim_{i.i.d.} \text{BG}(\theta)$ and $\lambda > 0$, define the set $I(\mathbf{a})$ in [\(G.1\)](#), write $\mathbf{s}_i = \sum_{\ell \neq i} \beta_{\ell} \mathbf{x}_{0\ell}$, then for every $\mathbf{a} \in \mathbb{S}^{p-1}$ and $i \in [n]$:*

$$n^{-1} \mathbb{E} \left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0] \right\|_2^2 = \theta \left[1 - \mathbb{E}_{\mathbf{s}_i} \text{erf}_{\beta_i}(\lambda, \mathbf{s}_i) + \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i) \right] \quad (\text{G.7})$$

Proof. Write \mathbf{x}_0 as \mathbf{x} . Observe that $\mathbf{y} * \check{\mathbf{a}} = \mathbf{x}_0 * \check{\boldsymbol{\beta}} = \sum_{\ell} \beta_{\ell} s_{-\ell}[\mathbf{x}_0]$. Thus for any $j \in [n]$ and $i \in [\pm p]$:

$$(\mathbf{y} * \check{\mathbf{a}})_{j-i} = \left(\beta_i s_{-i}[\mathbf{x}] + \sum_{\ell \neq i} \beta_{\ell} s_{-\ell}[\mathbf{x}] \right)_{j-i} = \beta_i \mathbf{x}_j + \sum_{\ell \neq i} \beta_{\ell} \mathbf{x}_{j+\ell-i} =: \beta_i \mathbf{x}_j + \mathbf{s}_j, \quad (\text{G.8})$$

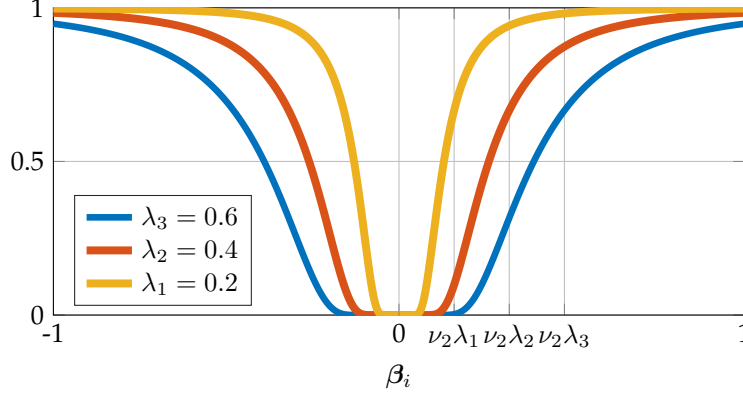


Figure 4: A numerical example for $\mathbb{E} \|\mathbf{P}_{I(\mathbf{a})} s_i[\mathbf{x}_0]\|_2^2$. We provide a figure to illustrate the expectation of $-\frac{1}{n\theta} \|\mathbf{P}_{I(\mathbf{a})} s_i[\mathbf{x}_0]\|_2^2$ when entries of \mathbf{x}_0 are $2p$ -separated, as a function plot of $\beta_i \rightarrow 1 - \text{erf}_{\beta_i}(\lambda, 0) + f_{\beta_i}(\lambda, 0)$ from (G.4) with different λ . When $|\beta_i| \approx \nu_2 \lambda$ where $\nu_2 = \sqrt{2/\pi}$, then the its function value is close to 0.5. If $|\beta_i|$ is much larger then λ its value grow to 1, implies there is a negative curvature at $s_i[\mathbf{a}_0]$ direction. Similarly if $|\beta_i|$ is much smaller then λ the function value is 0 thus the curvature is positive in $s_i[\mathbf{a}_0]$ direction.

where \mathbf{x}_j is independent of s_j , and both \mathbf{x}_j, s_j are symmetric and identically distributed for all $j \in [n]$. Rewrite the random variable using (G.1) as

$$\begin{aligned} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2^2 &= \left\| \mathbf{P}_{I(\mathbf{a})} \sum_{j \in [n]} (\mathbf{x}_{0j} \mathbf{e}_{j-i}) \right\|_2^2 = \sum_{j \in [n]} \mathbf{x}_{0j}^2 \mathbf{1}_{\{|\mathbf{y}^* \tilde{\mathbf{a}}|_{j-i} > \lambda\}} \\ &= \sum_{j \in [n]} \mathbf{x}_{0j}^2 \mathbf{1}_{\{|\beta_i \mathbf{x}_{0j} + s_j| > \lambda\}} \end{aligned}$$

Write $\mathbf{x} = \mathbf{g} \circ \omega$ as composition of Gaussian/Bernoulli r.v.s., the expectation has a simple form:

$$\mathbb{E} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2^2 = n\theta \cdot \mathbb{E} \mathbf{g}_0^2 \mathbf{1}_{\{|\beta_i \mathbf{g}_0 + s_0| > \lambda\}} = n\theta \cdot \mathbb{E} (1 - \text{erf}_{\beta_i}(\lambda, s_i) + f_{\beta_i}(\lambda, s_i))$$

where $s_i = \sum_{\ell \neq i} \mathbf{x}_{0\ell} \beta_\ell$ with $\mathbf{x}_{0\ell} \sim \text{i.i.d. BG}(\theta)$, yielding the claimed expression. \blacksquare

When the signal length of \mathbf{y} is sufficiently large, then i -th diagonal term for Hessian $\|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2^2$ will be close enough to its expected value.

Corollary G.3 (Large sample deviation of curvature). Suppose $\mathbf{x}_0 \sim \text{i.i.d. BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda / \sqrt{k}$ in φ_{ℓ^1} for some $c_\lambda > 1/5$, then there exists some numerical constant $C, c, \bar{c} > 0$, such that if $n \geq Cp^4 \theta^{-1} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - 3/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and every $i \in [n]$, we have:

$$\left| n^{-1} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2^2 - n^{-1} \mathbb{E} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2^2 \right| \leq c\theta/p \quad (\text{G.9})$$

Proof. See Appendix L.2. \blacksquare

The off-diagonal entries of Hessian in general are much smaller then the diagonal entries; however, it affects the region near sign shifts of \mathbf{a}_0 the most where we need to show strong convexity in the region. We provide an upper bound for off-diagonal entries in the vicinity of signed shifts. In these regions, only one entry of the correlations $|\beta_{(0)}|$ is large and the rest is small.

Lemma G.4 (Hessian off-diagonal term near solution). *Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Let $\lambda = c_\lambda/\sqrt{k}$ with $c_\lambda > 1/5$, then there exists some numerical constant $C, \bar{c} > 0$ such that if $n \geq C\theta^{-4} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - 4/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$, where $|\beta_{(1)}| \leq \frac{1}{4 \log \theta^{-1}} \lambda$ and every $i \neq j \in [\pm p] \setminus \{0\}$, we have*

$$|s_i[\mathbf{x}_0]^* | \mathbf{P}_{I(\mathbf{a})} | s_j[\mathbf{x}_0]| < 8n\theta^3 \quad (\text{G.10})$$

Proof. Write $\theta_{\log} = -1/\log \theta$ and \mathbf{x}_0 as $\mathbf{x} = \boldsymbol{\omega} \circ \mathbf{g}$. Wlog let β_0 be the largest correlation $\beta_{(0)}$. Define random variables $s' = \langle \boldsymbol{\beta}_{\tau \setminus \{0, i, j\}}, \mathbf{x}_{\tau \setminus \{0, i, j\}} \rangle$. Firstly via [Corollary E.7](#) we have $\mathbb{P}[|s'| > 0.4\lambda] \leq 2\theta$; also define $s = \langle \boldsymbol{\beta}_{\tau^c \setminus \{0, i, j\}}, \mathbf{x}_{\tau^c \setminus \{0, i, j\}} \rangle$, and base on [Corollary E.6](#) we have $\mathbb{P}[|s| > \lambda/10] \leq 2\theta$. Expand the $(-i, -j)$ -th cross term with $\theta < 0.1$ we have:

$$\begin{aligned} \mathbb{E} |s_{-i}[\mathbf{x}]^* | \mathbf{P}_{I(\mathbf{a})} | s_{-j}[\mathbf{x}]| &= \mathbb{E} \sum_{k \in [n]} |\mathbf{x}_{k+i} \mathbf{x}_{k+j}| \mathbf{1}_{\{|\beta_0 \mathbf{x}_k + \beta_i \mathbf{x}_{k+i} + \beta_j \mathbf{x}_{k+j} + s + s' > \lambda\}} \\ &= n\theta^2 \cdot \mathbb{E} |\mathbf{g}_i \mathbf{g}_j| \mathbf{1}_{\{|\beta_0 \mathbf{x}_0 + \beta_i \mathbf{g}_i + \beta_j \mathbf{g}_j + s + s' > \lambda\}} \\ &\leq n\theta^2 \cdot \mathbb{E} [|\mathbf{g}_i \mathbf{g}_j| (2\mathbf{1}_{\{|\beta_i \mathbf{g}_i| > \lambda/4\}} \\ &\quad + \mathbb{P}[\mathbf{x}_0 \neq 0] + \mathbb{P}[|s| > 0.1\lambda] + \mathbb{P}[|s'| > 0.4\lambda])] \\ &\leq n\theta^2 \cdot (\exp(-\log^2 \theta^{-1}) + \theta + 2\theta + 2\theta) \\ &\leq 6n\theta^3. \end{aligned} \quad (\text{G.11})$$

Write (G.10) as two summation of independent random variables with $t = j - i$ by separating sum into two sets J_{t1}, J_{t2} defined in (D.4) with both $|J_{t1}|, |J_{t2}| < n\theta^2$ with probability at least $1 - 2/n$ from [Lemma D.1](#)

$$\begin{aligned} \mathbb{E} |s_{-i}[\mathbf{x}]^* | \mathbf{P}_{I(\mathbf{a})} | s_{-j}[\mathbf{x}]| &= \sum_{(k-i) \in I(\mathbf{a})} |\mathbf{x}_k| |\mathbf{x}_{k+t}| \\ &= \sum_{(k-i) \in I(\mathbf{a}) \cap J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k+t}| + \sum_{(k-i) \in I(\mathbf{a}) \cap J_{t2}} |\mathbf{g}_k| |\mathbf{g}_{k+t}|, \end{aligned}$$

whose first summands can be upper bounded with high probability via Bernstein inequality [Lemma N.4](#) with $(\sigma^2, R) = (1, 1)$ and writes $\mathcal{C} := \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu)) \cap \left\{ \mathbf{a} \mid |\beta_{(1)}| \leq \frac{1}{4 \log \theta^{-1}} \lambda \right\}$, then we have

$$\begin{aligned} &\mathbb{P} \left[\max_{\substack{i \neq j \in [\pm p] \setminus \{0\} \\ \mathbf{a} \in \mathcal{C}}} \left(\sum_{(k-i) \in I(\mathbf{a}) \cap J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k+t}| - \mathbb{E} \sum_{(k-i) \in I(\mathbf{a}) \cap J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k+t}| \right) \geq n\theta^3 \right] \\ &\mathbb{P} \left[\max_{i \neq j \in [\pm p] \setminus \{0\}} \left(\sum_{(k-i) \in \cap J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k+t}| - \mathbb{E} \sum_{(k-i) \cap J_{t1}} |\mathbf{g}_k| |\mathbf{g}_{k+t}| \right) \geq n\theta^3 \right] \\ &\leq 4p^2 \cdot \exp\left(\frac{-n^2\theta^6}{2|J_{t1}| + 2n\theta^3}\right) \leq \exp\left(8 \log p - \frac{n^2\theta^6}{3n\theta^2}\right) \leq \exp\left(-\frac{n\theta^4}{10}\right) \leq \frac{1}{n} \end{aligned} \quad (\text{G.12})$$

when $n = C\theta^{-4} \log p$ with $C > 10^4$ and $\theta \log^2 \theta^{-1} \geq 1/p$. Thus for all $i \neq j \in [\pm p] \setminus \{0\}$ and \mathbf{a} satisfies our condition of lemma, from (G.11) and (G.12) we can conclude :

$$|s_{-i}[\mathbf{x}]^* | \mathbf{P}_{I(\mathbf{a})} | s_{-j}[\mathbf{x}]| \leq \sum_{I(\mathbf{a}) \cap J_{t1}} \mathbb{E} |\mathbf{g}_k| |\mathbf{g}_{k+t}| + \sum_{I(\mathbf{a}) \cap J_{t2}} \mathbb{E} |\mathbf{g}_k| |\mathbf{g}_{k+t}| + 2n\theta^3 \leq 8n\theta^3$$

which holds with probability at least $1 - 2/n - 2 \cdot 1/n = 1 - 4/n$ base on [Lemma D.1](#) and (G.12). \blacksquare

H Geometric relation between ρ and ℓ^1 -norm

In this section, we discuss how to ensure that the smooth sparsity surrogate ρ approximates $\|\cdot\|_1$ accurately enough that guarantees φ_ρ inherits the good properties of φ_{ℓ^1} . We prove several lemmas which allow us to transfer properties of φ_{ℓ^1} to φ_ρ . Our result does not pertain to the suggested pseudo-Huber surrogate $\rho(x)_i = \sqrt{x_i^2 + \delta^2}$ in the main script, and is general for a class of function class defined in [Definition H.2](#) that is smooth and well approximates ℓ^1 when the proper smoothing parameter δ is chosen from the result of [Lemma H.6](#). In particular we ask the regularizer $\rho_\delta(x)$ to be uniformly bounded to $|x|$ by $\delta/2$:

$$\forall x \in \mathbb{R}, \quad |\rho_\delta(x) - |x|| \leq \delta/2 \quad (\text{H.1})$$

then if $\delta \rightarrow 0$ we have for every \mathbf{a} near subspace,

$$\|\text{prox}_{\lambda\ell^1}[\check{\mathbf{a}} * \mathbf{y}] - \text{prox}_{\lambda\rho_\delta}[\check{\mathbf{a}} * \mathbf{y}]\|_2 \rightarrow 0, \quad (\text{H.2})$$

$$\|\nabla\varphi_{\ell^1}(\mathbf{a}) - \nabla\varphi_{\rho_\delta}(\mathbf{a})\|_2 \rightarrow 0, \quad (\text{H.3})$$

$$\|\tilde{\nabla}^2\varphi_{\ell^1}(\mathbf{a}) - \nabla^2\varphi_{\rho_\delta}(\mathbf{a})\|_2 \rightarrow 0. \quad (\text{H.4})$$

An example choices of eligible smooth sparse surrogate is demonstrated in [Table 1](#).

The marginal minimizer over \mathbf{x} in (??) can be expressed in terms of the proximal operator [\[BC11\]](#) of ρ at point $\check{\mathbf{a}} * \mathbf{y}$:

$$\text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}] = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \left\{ \lambda\rho(\mathbf{x}) + \frac{1}{2} \|\mathbf{x}\|_2^2 - \langle \mathbf{a} * \mathbf{x}, \mathbf{y} \rangle \right\}.$$

Plugging in, we obtain

$$\varphi_\rho(\mathbf{a}) = \lambda\rho(\text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}]) + \frac{1}{2} \|\check{\mathbf{a}} * \mathbf{y} - \text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}]\|_2^2 - \frac{1}{2} \|\check{\mathbf{a}} * \mathbf{y}\|_2^2 + \frac{1}{2} \|\mathbf{y}\|_2^2 \quad (\text{H.5})$$

The objective function $\varphi_\rho(\mathbf{a})$ is a differentiable function of \mathbf{a} . This can be seen, e.g., by noting that

$$\varphi_\rho(\mathbf{a}) = \epsilon(\lambda\rho)(\check{\mathbf{a}} * \mathbf{y}) - \frac{1}{2} \|\check{\mathbf{a}} * \mathbf{y}\|_2^2 + \frac{1}{2} \|\mathbf{y}\|_2^2, \quad (\text{H.6})$$

where $\epsilon(g)(\mathbf{z}) = g(\text{prox}_g(\mathbf{z})) + \frac{1}{2} \|\mathbf{z} - \text{prox}_g(\mathbf{z})\|_2^2$ is the *Moreau envelope* of a function g . The Moreau envelope is differentiable:

Fact H.1 (Derivative of Moreau envelope, [\[BC11\]](#), Prop.12.29). *Let f be a proper lower semicontinuous convex function and $\lambda > 0$ then the Moreau envelope $\epsilon(\lambda f)(\mathbf{z}) = \lambda f(\text{prox}_{\lambda f}[\mathbf{z}]) + \frac{1}{2} \|\mathbf{z} - \text{prox}_{\lambda f}[\mathbf{z}]\|_2^2$ is Fréchet differentiable with $\nabla\epsilon(\lambda f)(\mathbf{z}) = \mathbf{z} - \text{prox}_{\lambda f}[\mathbf{z}]$.*

Furthermore, φ_ρ is twice differentiable whenever $\text{prox}_{\lambda\rho}$ is differentiable. In this case, the (Euclidean) gradient and hessian of φ_ρ are given by

$$\nabla\varphi_\rho(\mathbf{a}) = -\boldsymbol{\iota}^* \check{\mathbf{C}}_{\mathbf{y}} \text{prox}_{\lambda\rho}[\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a}], \quad (\text{H.7})$$

$$\nabla^2\varphi_\rho(\mathbf{a}) = -\boldsymbol{\iota}^* \check{\mathbf{C}}_{\mathbf{y}} \nabla \text{prox}_{\lambda\rho}[\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a}] \check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota}. \quad (\text{H.8})$$

The Riemannian gradient and hessian over \mathbb{S}^{p-1} are

$$\text{grad}[\varphi_\rho](\mathbf{a}) = -\mathbf{P}_{\mathbf{a}^\perp} \boldsymbol{\iota}^* \check{\mathbf{C}}_{\mathbf{y}} \text{prox}_{\lambda\rho}[\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a}], \quad (\text{H.9})$$

$$\text{Hess}[\varphi_\rho](\mathbf{a}) = -\mathbf{P}_{\mathbf{a}^\perp} \left(\boldsymbol{\iota}^* \check{\mathbf{C}}_{\mathbf{y}} \nabla \text{prox}_{\lambda\rho}[\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a}] \check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} - \langle \nabla\varphi_\rho(\mathbf{a}), \mathbf{a} \rangle \mathbf{I} \right) \mathbf{P}_{\mathbf{a}^\perp}. \quad (\text{H.10})$$

Our analysis accommodates any sufficiently accurate smooth approximation ρ to the ℓ^1 function. The requisite sense of approximation is captured in the following definition:

| Surrogate class | $\rho_i(x)$ | $\nabla\rho_i(x)$ | $\nabla^2\rho_i(x)$ |
|-----------------------|---|---|--|
| Log hyperbolic cosine | $\frac{\delta}{2} \log(e^{2x/\delta} + e^{-2x/\delta})$ | $\frac{e^{4x/\delta} - 1}{e^{4x/\delta} + 1}$ | $\frac{4e^{4x/\delta}}{\delta(e^{4x/\delta} + 1)^2}$ |
| Pseudo Huber | $\sqrt{x^2 + \delta^2}$ | $\frac{x}{\sqrt{x^2 + \delta^2}}$ | $\frac{\delta^2}{(x^2 + \delta^2)^{3/2}}$ |
| Gaussian convolution | $\int x - t f_\delta(t) dt$ | $\text{erf}(x/\sqrt{2}\delta)$ | $2f_\delta(x)$ |

Table 1: Classes of smooth sparse surrogate ρ and how to set its parameter. Three common classes are listed with parameter δ to tune the smoothness. All the listed functions are greater than $|x|$ pointwise and has largest distance to $|x|$ at origin where $\rho(0) - |x| \leq \delta$, satisfies the condition (H.11). Also its second order derivatives $\nabla^2\rho_i(x)$ are monotone decreasing w.r.t. $|x|$, hence are certified to be eligible δ -smoothed ℓ^1 surrogates.

Definition H.2 (δ -smoothed ℓ^1 function). *We call an additively separable function $\rho(\mathbf{x}) = \sum_{i=1}^n \rho_i(\mathbf{x}_i) : \mathbb{R}^n \rightarrow \mathbb{R}$, a δ -smoothed ℓ^1 function with $\delta > 0$ if for each $i \in [n]$, ρ_i is even, convex, twice differentiable and $\nabla^2\rho_i(x)$ being monotone decreasing w.r.t. $|x|$, where, there exists some constant c , such that for all $x \in \mathbb{R}$:*

$$|\rho_i(x) - |x| + c| \leq \delta/2 \quad (\text{H.11})$$

The proximal operator of the ℓ^1 norm is the entrywise soft thresholding function \mathcal{S}_λ ; the proximal operator associated to a smoothed ℓ^1 function turns out to be a differentiable approximation to \mathcal{S}_λ . In particular, we will show that it approximates \mathcal{S}_λ in the following sense:

Definition H.3 ($\sqrt{\delta}$ -smoothed soft threshold). *An odd function $\mathcal{S}_\lambda^\delta[\cdot] : \mathbb{R} \rightarrow \mathbb{R}$ is a $\sqrt{\delta}$ -smoothed soft thresholding function with parameter $\delta > 0$ if it is a strictly monotone odd function and is differentiable everywhere, whose function value satisfies*

$$0 \leq \text{sign}(z) (\mathcal{S}_\lambda^\delta[z] - \mathcal{S}_\lambda[z]) \leq \sqrt{\lambda\delta}, \quad \forall z \in \mathbb{R} \quad (\text{H.12})$$

and its derivative satisfies for any given $B \in (0, \lambda)$:

$$|\nabla\mathcal{S}_\lambda^\delta[z] - \nabla\mathcal{S}_\lambda[z]| \leq \sqrt{\lambda\delta}/B, \quad ||z| - \lambda| \geq B. \quad (\text{H.13})$$

If ρ is a δ -smooth ℓ^1 function, then for all $i \in [n]$, we have that $\text{prox}_{\lambda\rho}[z]_i$ is a $\sqrt{\delta}$ -smoothed soft threshold function of z_i . This can be proven with the following lemma:

Lemma H.4 (Proximal operator for smoothed ℓ^1). *Suppose ρ is a δ -smoothed ℓ^1 function, then $z_i \mapsto \text{prox}_{\lambda\rho}[z]_i$ is a $\sqrt{\delta}$ -smoothed soft threshold function.*

Proof. We know that

$$\mathbf{x}_z := \text{prox}_{\lambda\rho}[\mathbf{z}] = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \lambda\rho(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2. \quad (\text{H.14})$$

This optimization problem is strongly convex, and so the minimizer \mathbf{x}_z is unique. Using the stationarity condition and since ρ is separable, for all $i \in [n]$, we have $\lambda\nabla\rho_i(\mathbf{x}_{zi}) + \mathbf{x}_{zi} - \mathbf{z}_i = 0$, implies

$$\mathbf{x}_{zi} = (\text{Id} + \lambda\nabla\rho_i)^{-1}(\mathbf{z}_i). \quad (\text{H.15})$$

Since ρ_i is convex and even, $\nabla\rho_i$ is monotone increasing and odd. By inverse function theorem, we know that strict monotonicity and differentiability of $\text{Id} + \lambda\nabla\rho_i$ implies its inverse is differentiable and is a strictly monotone increasing odd function. Furthermore, it implies $\nabla\mathbf{x}_{z_i}$ has the form

$$\nabla\mathbf{x}_{z_i} = \nabla_i(\text{Id} + \lambda\nabla\rho_i)^{-1}(z_i) = \frac{1}{\lambda\nabla^2\rho_i(\mathbf{x}_{z_i}) + 1} < 1. \quad (\text{H.16})$$

Notice that since $\nabla^2\rho_i(x)$ is monotone decreasing when $x \geq 0$, hence $\nabla\mathbf{x}_{z_i}$ is monotone increasing in $z_i \geq 0$.

Now we are left to show that (H.12) and (H.13) hold, and since $\text{prox}_{\lambda\rho}[\cdot]_i$ is an odd function it suffices to consider the case when the input vector z_i is nonnegative. Firstly, via convexity and entrywise bounded difference $|\rho_i(x) - |x|| \leq \delta/2$ we are going to show

$$|\nabla\rho_i(x)| \leq 1 \quad \forall x \in \mathbb{R}, \quad \nabla\rho_i(x) \geq 1 - \sqrt{\delta/\lambda} \quad \forall x \geq \sqrt{\lambda\delta}. \quad (\text{H.17})$$

Consider a positive x with $\nabla\rho_i(x) > 1 + \varepsilon$ for some $\varepsilon > 0$, by convexity if $\tilde{x} > x$ then $\nabla\rho_i(\tilde{x}) > 1 + \varepsilon$, hence

$$\rho_i(x + \delta/\varepsilon) \geq \rho_i(x) + \nabla\rho_i(x) \cdot (\delta/\varepsilon) > x - \delta/2 + (1 + \varepsilon) \cdot (\delta/\varepsilon) = (x + \delta/\varepsilon) + \delta/2,$$

contradicts the boundedness condition. Secondly, use mean value theorem we know for all $x \geq \sqrt{\lambda\delta}$:

$$\nabla\rho_i(x) \geq \frac{\rho_i(\sqrt{\lambda\delta}) - \rho_i(0)}{\sqrt{\lambda\delta} - 0} \geq \frac{(\sqrt{\lambda\delta} - \delta/2) - (0 + \delta/2)}{\sqrt{\lambda\delta} - 0} \geq 1 - \sqrt{\frac{\delta}{\lambda}}.$$

To prove (H.12), when $0 \leq z_i \leq \lambda$, then $\mathcal{S}_\lambda[z_i] = 0$ and $\mathbf{x}_{z_i} \leq \sqrt{\lambda\delta}$ since if $\mathbf{x}_{z_i} > \sqrt{\lambda\delta}$, by (H.17):

$$\lambda\nabla\rho_i(\mathbf{x}_{z_i}) + \mathbf{x}_{z_i} > \lambda(1 - \sqrt{\delta/\lambda}) + \sqrt{\lambda\delta} = \lambda \geq z_i$$

then \mathbf{x}_{z_i} violate the stationary condition in (H.15), resulting $0 \leq \mathbf{x}_{z_i} - \mathcal{S}_\lambda[z_i] \leq \sqrt{\lambda\delta}$ whenever $0 \leq z_i \leq \lambda$. Likewise in the case of $z_i \geq \lambda$ where $\mathcal{S}_\lambda[z_i] = z_i - \lambda$, (H.17) provides:

$$\begin{cases} \forall \mathbf{x}_{z_i} > z_i - \lambda + \sqrt{\lambda\delta}, & \lambda\nabla\rho_i(\mathbf{x}_{z_i}) + \mathbf{x}_{z_i} > \lambda(1 - \sqrt{\delta/\lambda}) + z_i - \lambda + \sqrt{\lambda\delta} = z_i \\ \forall \mathbf{x}_{z_i} < z_i - \lambda, & \lambda\nabla\rho_i(\mathbf{x}_{z_i}) + \mathbf{x}_{z_i} < \lambda + z_i - \lambda = z_i \end{cases}$$

again violates (H.15) and therefore (H.12) holds for all $z_i \in \mathbb{R}$.

Lastly (H.13) is a direct result of (H.12). For all $z_i \leq \lambda - B$, recall that $\nabla\mathbf{x}_{z_i}$ is monotone increasing in z_i :

$$\nabla\mathbf{x}_{z_i} \leq \min_{y \in [\lambda - B, \lambda]} \nabla\mathbf{x}_{y_i} \leq \frac{\mathbf{x}_{\lambda_i} - \mathbf{x}_{(\lambda - B)_i}}{\lambda - (\lambda - B)} \leq \frac{(\sqrt{\lambda\delta} + \mathcal{S}_\lambda[\lambda]) - \mathcal{S}_\lambda[\lambda - B]}{B} = \frac{\sqrt{\lambda\delta}}{B};$$

and similarly for all $z_i > \lambda + B$:

$$\nabla\mathbf{x}_{z_i} \geq \max_{y \in [\lambda, \lambda + B]} \nabla\mathbf{x}_{y_i} \geq \frac{\mathbf{x}_{(\lambda + B)_i} - \mathbf{x}_{\lambda_i}}{(\lambda + B) - \lambda} \geq \frac{\mathcal{S}_\lambda[\lambda + B] - (\mathcal{S}_\lambda[\lambda] + \sqrt{\lambda\delta})}{B} = 1 - \frac{\sqrt{\lambda\delta}}{B},$$

implies (H.13) holds. ■

Based on (H.9)-(H.10) and denote $\check{\mathbf{C}}_{\mathbf{y}}\boldsymbol{\iota}\mathbf{a} = \check{\mathbf{a}} * \mathbf{y}$, the only differences of Riemannian gradient and Hessian between φ_ρ and φ_{ℓ^1} comes from the difference of $\text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}]$ and $\text{prox}_{\lambda\|\cdot\|_1}[\check{\mathbf{a}} * \mathbf{y}]$. Thus for the purpose of obtaining good geometric approximation of φ_ρ with that of objective φ_{ℓ^1} , we may apply both Definition H.3 and Lemma H.4, together suggest if ρ is a δ -smoothed ℓ^1 function, then the i -th entry of $\text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}]$ will be $\sqrt{\lambda\delta}$ -close to the authentic soft thresholding function $\mathcal{S}_\lambda[\check{\mathbf{a}} * \mathbf{y}]_i$, and its gradient $\nabla \text{prox}_{\lambda\rho}[\check{\mathbf{a}} * \mathbf{y}]$ is $\sqrt{\lambda\delta}/B$ -close to $\nabla\mathcal{S}_\lambda[\check{\mathbf{a}} * \mathbf{y}]$ as long as $(\check{\mathbf{a}} * \mathbf{y})_i$ is not close to $\pm\lambda$ by distance B .

Firstly, we will show by utilizing the random structure of \mathbf{y} , such that with high probability, only a fraction of entries of $\check{\mathbf{a}} * \mathbf{y}$ will be close to $\pm\lambda$.

Lemma H.5 (Gradients discontinuity entries). For each $\mathbf{a} \in \mathbb{S}^{p-1}$, let

$$J_B(\mathbf{a}) := \left\{ i \mid \left(\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota} \mathbf{a} \right)_i \in [-\lambda - B, -\lambda + B] \cup [\lambda - B, \lambda + B] \right\}. \quad (\text{H.18})$$

Suppose the subspace dimension is at most k and signal \mathbf{y} satisfies [Definition E.1](#). Let $\lambda = c_\lambda / \sqrt{k}$ and $B \leq c' \lambda \theta^2 / p \log n$ for some $c_\lambda, c' \in (0, 1)$, then there is a numerical constant $C > 0$ such that if $n \geq Cp^5 \theta^{-2} \log p$, then with probability at least $1 - 3/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$, we have

$$|J_B(\mathbf{a})| \leq \frac{24c'n\theta^2}{p \log n} \quad (\text{H.19})$$

Proof. See [Appendix L.3](#). ■

The geometric approximation between φ_{ℓ^1} and φ_ρ necessarily consists of three parts: the gradient, the Hessian, and the coefficients. Here we conclude the approximation result with the following lemma:

Lemma H.6 (φ_{ℓ^1} approximates φ_ρ). Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Let $\rho \in \mathbb{R}^n \rightarrow \mathbb{R}$ be a δ -smoothed ℓ^1 function with

$$\lambda = \frac{c_\lambda}{\sqrt{k}}, \quad \delta \leq \frac{c'^4 \theta^8}{p^2 \log^2 n} \lambda \quad (\text{H.20})$$

with some $c', c_\lambda \in (0, 1)$, then there is a numerical constant $C, \bar{c} > 0$ such that if $n > Cp^5 \theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - 10/n$, the following statements hold simultaneously for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$:

(1). The coefficients has norm difference

$$\left\| \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \text{prox}_{\lambda \ell^1}[\check{\mathbf{a}} * \mathbf{y}] - \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \text{prox}_{\lambda \rho}[\check{\mathbf{a}} * \mathbf{y}] \right\|_2 \leq c'n\theta^4. \quad (\text{H.21})$$

(2). The gradient has norm difference

$$\|\nabla \varphi_{\ell^1}(\mathbf{a}) - \nabla \varphi_\rho(\mathbf{a})\|_2 \leq c'n\theta^4. \quad (\text{H.22})$$

(3). The (pesudo) Riemmannian curvature difference is bounded in all directions $\mathbf{v} \in \mathbb{S}^{p-1}$ via

$$\forall \mathbf{v} \in \mathbb{S}^{p-1}, \quad \left| \mathbf{v}^* \left(\widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) - \text{Hess}[\varphi_\rho](\mathbf{a}) \right) \mathbf{v} \right| \leq 200c'n\theta^2. \quad (\text{H.23})$$

Proof. 1. (Coefficients) From [Lemma H.4](#), the proximal δ -smoothed ℓ^1 function satisfies

$$|\mathcal{S}_\lambda[\check{\mathbf{a}} * \mathbf{y}] - \mathcal{S}_\lambda^\delta[\check{\mathbf{a}} * \mathbf{y}]|_j < \sqrt{\lambda \delta} \quad \forall j \in [n].$$

Since the support of coefficient vectors are contained in $[\pm p]$, using simple norm inequality:

$$\left\| \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda[\check{\mathbf{a}} * \mathbf{y}] - \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda^\delta[\check{\mathbf{a}} * \mathbf{y}] \right\|_2 \leq \sqrt{\lambda \delta n} \cdot \left\| \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \right\|_2. \quad (\text{H.24})$$

Apply [Lemma D.5](#) by replacing \mathbf{a}_0 with standard basis \mathbf{e}_0 and extend support of $\boldsymbol{\iota}$ to $\boldsymbol{\iota}_{[\pm p]}$, notice that in this case we have $\mu = 0$. Condition on the event

$$\left\| \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \right\|_2 \leq \left\| \boldsymbol{\iota}_{[\pm p]}^* \check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{C}_{\mathbf{e}_0}^* \right\|_2 \leq \sqrt{3(1 + 2\mu p)n\theta} \leq \sqrt{3n\theta},$$

and we gain

$$(\text{H.24}) \leq \sqrt{\lambda \delta n} \cdot \sqrt{3n\theta} \leq n\sqrt{3\lambda \theta \delta} \leq c'n\theta^4.$$

2. (Gradient) From definition of Riemannian gradient (H.9) and apply similar norm bound of (H.24), and condition on the following events of Lemma D.5 holds, obtain

$$\|\nabla\varphi_{\ell^1}(\mathbf{a}) - \nabla\varphi_{\rho}(\mathbf{a})\|_2 \leq \sqrt{\lambda\delta n} \cdot \|\iota^* \check{\mathbf{C}}_{\mathbf{y}}\|_2 \leq n\sqrt{3\lambda\theta(1+\mu p)\delta} \leq c'n\theta^4. \quad (\text{H.25})$$

3. (Hessian) For every realization of $J_B(\mathbf{a})$ from $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_{\tau}, \gamma(c_{\mu}))$, base on Lemma H.5, condition on the event such that

$$B \leq \frac{c'\lambda\theta^2}{p \log n}, \quad |J| \leq \frac{24c'n\theta^2}{p \log n}; \quad (\text{H.26})$$

and rewrite $J_B(\mathbf{a})$ as J . Also condition on the event using Lemma D.5 and $(1+\mu p)\theta \log \theta^{-1} < 1$

$$\|\iota^* \check{\mathbf{C}}_{\mathbf{y}}\|_2 \leq \sqrt{3n}, \quad \|\iota^* \check{\mathbf{C}}_{\mathbf{y}} \mathbf{P}_J\|_2 \leq \sqrt{8|J|p \log n}, \quad (\text{H.27})$$

then the difference of Hessian (H.10), in direction $\mathbf{v} \in \mathbb{S}^{p-1}$ can be bounded as

$$\begin{aligned} & \left| \mathbf{v}^* \left(\widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) - \text{Hess}[\varphi_{\rho}](\mathbf{a}) \right) \mathbf{v} \right| \\ & \leq \left| \mathbf{v}^* \iota^* \check{\mathbf{C}}_{\mathbf{y}} \left(\mathbf{P}_{I(\mathbf{a})} - \text{diag} \left[\nabla \mathcal{S}_{\lambda}^{\delta} \left[\check{\mathbf{C}}_{\mathbf{y}} \iota \mathbf{a} \right] \right] \right) \check{\mathbf{C}}_{\mathbf{y}} \iota \mathbf{v} \right| + \|\nabla\varphi_{\ell^1}(\mathbf{a}) - \nabla\varphi_{\rho}(\mathbf{a})\|_2 \end{aligned} \quad (\text{H.28})$$

where $I(\mathbf{a})$ is defined in (G.1). Let $\mathbf{D} = \mathbf{P}_{I(\mathbf{a})} - \text{diag} \left[\nabla \mathcal{S}_{\lambda}^{\delta} \left[\check{\mathbf{C}}_{\mathbf{y}} \iota \mathbf{a} \right] \right]$ and notice that \mathbf{D} is a diagonal matrix, which suggests (H.28) can be decomposed using

$$(\mathbf{P}_J + \mathbf{P}_{J^c}) \mathbf{D} (\mathbf{P}_J + \mathbf{P}_{J^c}) = \mathbf{P}_J \mathbf{D} \mathbf{P}_J + \mathbf{P}_{J^c} \mathbf{D} \mathbf{P}_{J^c},$$

where, from with property of $\sqrt{\delta}$ -smoothed ℓ^1 function Lemma H.4:

$$\max_j |\mathbf{P}_J \mathbf{D} \mathbf{P}_J|_{jj} \leq 1, \quad \max_j |\mathbf{P}_{J^c} \mathbf{D} \mathbf{P}_{J^c}|_{jj} \leq \sqrt{\lambda\delta}/B.$$

Finally, once again apply δ bound from (H.20) and bounds for $B, |J|, \mathbf{y}$ from (H.26)-(H.27), we gain

$$\begin{aligned} (\text{H.28}) & \leq \left\| \iota^* \check{\mathbf{C}}_{\mathbf{y}} \mathbf{P}_J \right\|_2^2 + \frac{\sqrt{\lambda\delta}}{B} \left\| \iota^* \check{\mathbf{C}}_{\mathbf{y}} \right\|_2^2 + \|\nabla\varphi_{\ell^1}(\mathbf{a}) - \nabla\varphi_{\rho}(\mathbf{a})\|_2 \\ & \leq 8|J|p \log n + \frac{3n\sqrt{\lambda\delta}}{B} + c'n\theta^2 \\ & \leq 8 \cdot \frac{24c'n\theta^2}{p \log n} \cdot p \log n + \frac{3n(c'^4\lambda^2\theta^8/p^2 \log^2 n)^{1/2}}{c'\lambda\theta^2/p \log p} + c'n\theta^2 \\ & \leq 200c'n\theta^2, \end{aligned}$$

where all above result holds with probability at least $1 - 10/n$ from Lemma H.5 and Lemma D.5. \blacksquare

I Analysis of geometry

In this section we prove major geometrical result in Theorem B.1. This lemma consists of three parts of geometry of φ_{ρ} ; including the negative curvature region Corollary I.2, large gradient region Corollary I.4, strong convexity region near shift Corollary I.6, and retraction to subspace Corollary I.8, which are respectively base on geometric properties of φ_{ℓ^1} in Lemma I.1, Lemma I.3, Lemma I.5 and Lemma I.7. We will handle each individual region in the following subsections. To shed light on the technical detail of the proof, we will begin with two figures for illustration of a toy example, which demonstrate the geometry near a two dimension solution subspace $\mathcal{S}_{\{i,j\}}$, as follows:

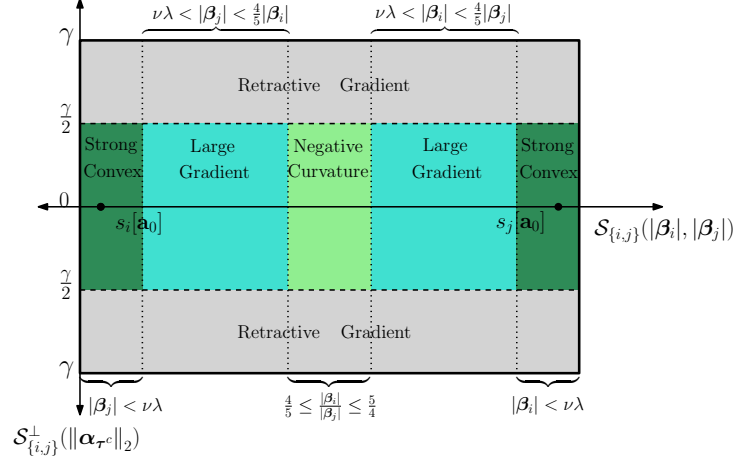


Figure 5: The top view of geometry over subspace $\mathcal{S}_{\{i,j\}}$. We display the geometric properties in the neighborhood of subspace $\mathcal{S}_{\{i,j\}}$ (horizontal axis) which contains the solutions $s_i[\mathbf{a}_0]$ and $s_j[\mathbf{a}_0]$. When \mathbf{a} lies near middle of two shifts (light green region) such that $|\beta_i| \approx |\beta_j|$, then there exists a negative curvature direction in subspace $\mathcal{S}_{\{i,j\}}$. When \mathbf{a} leans closer to one of the shifts $s_i[\mathbf{a}_0]$ (blue green region), its negative gradient direction points at that nearest shift. When \mathbf{a} is in the neighborhood of the shift $s_i[\mathbf{a}_0]$ (dark green region) such that $|\beta_i| \ll \lambda$, it will be strongly convex at \mathbf{a} , and the unique minimizer within the convex region will be close to $s_i[\mathbf{a}_0]$. Finally, the negative gradient will be pointing back toward the subspace $\mathcal{S}_{\{i,j\}}$ if near boundary (grey region).

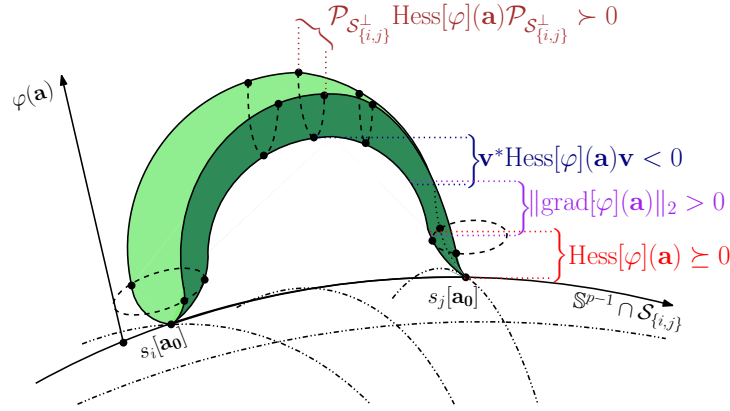


Figure 6: The side view of geometry of subspace $\mathcal{S}_{\{i,j\}}$ on sphere. We illustrate the geometry of $\mathcal{S}_{\{i,j\}}$ over the sphere, in which the properties of the three regions are denoted. In negative curvature region, there exists a direction \mathbf{v} such that $\mathbf{v}^* \text{Hess}[\varphi](\mathbf{a}) \mathbf{v}$ is negative. In large gradient region, the norm of Riemannian gradient $\|\text{grad}[\varphi](\mathbf{a})\|_2$ will be strictly greater than 0 and pointing at the nearest shift. Finally there is a convex region near all shifts such that $\text{Hess}[\varphi](\mathbf{a})$ is positive semidefinite.

I.1 Negative curvature

For any $\mathbf{a} \in \mathbb{S}^{p-1}$ near the subspace \mathcal{S}_τ such that the entries of leading correlation vector $\beta_{(0)}, \beta_{(1)}$ have balanced magnitude, the Hessian of $\varphi_\rho(\mathbf{a})$ exhibits negative curvature in the span of $s_{(0)}[\mathbf{a}_0], s_{(1)}[\mathbf{a}_0]$. We will first demonstrate the pseudo negative curvature of φ_{ℓ^1} in [Lemma I.1](#), then show φ_ρ approximates φ_{ℓ^1} in terms of Hessian in [Corollary I.2](#) when ρ is properly defined as in [Appendix H](#).

Lemma I.1 (Negative curvature for φ_{ℓ^1}). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Set $\lambda = c_\lambda / \sqrt{k}$ in φ_{ℓ^1} with $c_\lambda \in [\frac{1}{5}, \frac{1}{4}]$. There exist numerical*

constants $C, c, c', \bar{c} > 0$ such that if $n > Cp^5\theta^{-2} \log p$, and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$ the following holds at every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ satisfying $|\beta_{(1)}| \geq \frac{4}{5} |\beta_{(0)}|$: for $\mathbf{v} \in \mathcal{S}_{\{(0),(1)\}} \cap \mathbb{S}^{p-1} \cap \mathbf{a}^\perp$,

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) \mathbf{v} \leq -cn\theta\lambda. \quad (\text{I.1})$$

Proof. First of all the regional condition $\left| \frac{\beta_{(0)}}{\beta_{(1)}} \right| \leq \frac{5}{4}$ provides a two side bound for the two leading β 's

$$0.79 \geq \frac{|\beta_{(0)}|}{\sqrt{\beta_{(0)}^2 + \beta_{(1)}^2}} \|\beta_\tau\|_2 \geq |\beta_{(0)}| \geq |\beta_{(1)}| \geq \frac{4}{5} |\beta_{(0)}| \geq \frac{4}{5} \cdot \frac{\|\beta_\tau\|_2}{\sqrt{|\tau|}} \geq \frac{0.79}{\sqrt{|\tau|}} \quad (\text{I.2})$$

Set $J = \{(0), (1)\}$, choose $\mathbf{v} = \iota^* \mathbf{C}_{\mathbf{a}_0} \iota_J \gamma$ with $\|\mathbf{v}\|_2 = 1$ then $\left| \|\gamma\|_2^2 - 1 \right| \leq \mu$. There exists such \mathbf{v} satisfies condition above with $\mathbf{a} \perp \mathbf{v}$ by choosing γ as

$$\mathbf{a}^* \mathbf{v} = \mathbf{a}^* \iota^* \mathbf{C}_{\mathbf{a}_0} \iota_J \gamma = \gamma_{(0)} \beta_{(0)} + \gamma_{(1)} \beta_{(1)} = 0,$$

hence $\left| \frac{\gamma_{(1)}}{\gamma_{(0)}} \right| = \left| \frac{\beta_{(0)}}{\beta_{(1)}} \right| \leq \frac{5}{4}$. This implies $\gamma_{(0)}^2 \geq \frac{16}{25} \gamma_{(1)}^2 \geq \frac{16}{25} (1 - \mu - \gamma_{(0)}^2)$ where $\mu \leq \frac{c_\mu}{4} \leq \frac{1}{100}$, it gives the lower bound of $\gamma_{(0)}$ as

$$\gamma_{(0)}^2 \geq \frac{(1 - \mu) \cdot 16}{25 + 16} \geq 0.385 \quad (\text{I.3})$$

1. (Expand the Hessian) The (pseudo) curvature along direction \mathbf{v} is written as

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) \mathbf{v} = \mathbf{v}^* \widetilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v} - \langle \nabla \varphi_{\ell^1}(\mathbf{a}), \mathbf{a} \rangle = -\gamma^* \iota_J^* \mathbf{M} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{M} \iota_J \gamma + \beta^* \chi[\beta] \quad (\text{I.4})$$

expand the first term of (I.4) we obtain

$$\begin{aligned} & -\gamma^* \iota_J^* \mathbf{M} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{M} \iota_J \gamma \\ &= -\gamma^* \iota_J^* \mathbf{M} (\mathbf{P}_{(0)} + \mathbf{P}_{(1)} + \mathbf{P}_{J^c}) \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} (\mathbf{P}_{(0)} + \mathbf{P}_{(1)} + \mathbf{P}_{J^c}) \mathbf{M} \iota_J \gamma \\ &\leq -\sum_{i \in J} \left\| \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_i \right\|_2^2 (e_i^* \mathbf{M} \iota_J \gamma)^2 + 2 \sum_{\substack{(i,j) \in \{J, J^c\} \\ (i,j) \neq ((0),(1))}} \left| e_i^* \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| |(e_i^* \mathbf{M} \iota_J \gamma) (e_j^* \mathbf{M} \iota_J \gamma)| \\ &\leq -\sum_{i \in J} \left\| \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_i \right\|_2^2 (|\gamma_i| - \mu)^2 \\ &\quad + 2 \max_{i \neq j \in [\pm p]} \left| e_i^* \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| (\|\iota_J^* \mathbf{M} \iota_J \gamma\|_1 \|\iota_{J^c}^* \mathbf{M} \iota_J \gamma\|_1 + (|\gamma_{(0)}| + \mu) (|\gamma_{(1)}| + \mu)) \end{aligned} \quad (\text{I.5})$$

Consider the following events

$$\begin{cases} \mathcal{E}_{\text{cross}} := \left\{ \forall \mathbf{a} \in \mathbb{S}^{p-1}, \max_{i \neq j \in [\pm p]} \left| e_i^* \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| < 4n\theta^2 \right\} \\ \mathcal{E}_{\text{ncurv}} := \left\{ \forall \mathbf{a} \in \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu)), \min_{i \in J} \left\| \mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}] \right\|_2^2 \geq n\theta (1 - \mathbb{E}_{\mathbf{s}_i}(\lambda, \mathbf{s}_i) + \mathbb{E}_{\mathbf{s}_i}(\lambda, \mathbf{s}_i)) - \frac{c_\mu n\theta}{p} \right\} \end{cases}, \quad (\text{I.6})$$

and from Lemma E.4 we know

$$\|\iota_J^* \mathbf{M} \iota_J \gamma\|_1 \leq \|\gamma\|_1 + 2\mu \leq 1.5, \quad \|\iota_{J^c}^* \mathbf{M} \iota_J \gamma\|_1 \leq \mu p \|\gamma\|_1 \leq 1.5\mu p,$$

on the event $\mathcal{E}_{\text{cross}} \cap \mathcal{E}_{\text{ncurv}}$, we have

$$-\gamma^* \iota_J^* \mathbf{M} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{M} \iota_J \gamma$$

$$\leq \underbrace{-n\theta \cdot \sum_{i \in J} (|\gamma_i| - \mu)^2 (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i) + \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i))}_{g_1(\beta)} + (18\mu p + 8)n\theta^2 + \frac{2c_\mu n\theta}{\sqrt{|\tau|}} \quad (I.7)$$

Meanwhile, for the latter term of (I.4), consider the following event $\mathcal{E}_{\bar{\chi}}$ where we write $\sigma_i = \operatorname{sign}(\beta_i)$ as:

$$\mathcal{E}_{\bar{\chi}} := \left\{ \sigma_i \chi[\beta]_i \leq \begin{cases} n\theta \cdot |\beta_i| (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i)) + \frac{c_\mu n\theta}{p}, & \forall i \in \tau \\ n\theta \cdot |\beta_i| 4\theta |\tau| + \frac{c_\mu n\theta}{p}, & \forall i \in \tau^c \end{cases} \right\}, \quad (I.8)$$

and use both $\|\beta\|_1 \leq \frac{c_\mu p}{\sqrt{|\tau|}}$, $\|\beta_{\tau^c}\|_2^2 \leq \frac{c_\mu}{\theta |\tau|^2}$. On this event we have

$$\begin{aligned} \beta^* \chi[\beta] &\leq n\theta \cdot \sum_{i \in \tau} \beta_i^2 (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i)) + 4n\theta^2 |\tau| \|\beta_{\tau^c}\|_2^2 + \frac{c_\mu n\theta}{p} \|\beta\|_1 \\ &\leq n\theta \cdot \underbrace{\sum_{i \in \tau} \beta_i^2 (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i))}_{g_2(\beta)} + \frac{5c_\mu n\theta}{\sqrt{|\tau|}}. \end{aligned} \quad (I.9)$$

2. (Lower bound $\mathbb{E} f_{\beta_i}$) Combine the first term from each of the (I.7) and (I.9). Use $\mu \leq c_\mu \leq \frac{1}{300}$ and (I.3) to obtain $(|\gamma_{(0)}| - \mu)^2 > 0.38$, we have

$$\begin{aligned} \frac{1}{n\theta} (g_1(\beta) + g_2(\beta)) &\leq - \sum_{i \in J} \left[(|\gamma_i| - \mu)^2 - \beta_i^2 \right] (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i)) \\ &\quad + \sum_{i \in \tau \setminus J} \beta_i^2 (1 - \mathbb{E}_{\mathbf{s}_i} \operatorname{erf}_{\beta_i}(\lambda, \mathbf{s}_i)) - 0.38 \sum_{i \in J} \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i), \end{aligned} \quad (I.10)$$

now use Taylor expansion² for f_{β_i} , and apply the upper bound where $\mathbb{E} \mathbf{s}_i^2 \leq \theta \|\beta\|_2^2 \leq \theta \left(1 + \frac{c_\mu}{\sqrt{|\tau|}} + \frac{c_\mu}{\theta |\tau|^2} \right) \leq \frac{3c_\mu}{|\tau|}$,

$$\begin{aligned} \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i) &\geq \mathbb{E}_{\mathbf{s}_i} \frac{1}{\sqrt{2\pi}} \cdot \left(\frac{2\lambda}{|\beta_i|} - \frac{\lambda^3}{|\beta_i|^3} \left(1 + \frac{3\mathbf{s}_i^2}{\lambda^2} \right) \right) \\ &\geq \frac{1}{\sqrt{2\pi}} \cdot \underbrace{\left(\frac{2\lambda}{|\beta_i|} - \frac{1}{|\beta_i|^3} \left(\lambda^3 + \frac{9c_\mu \lambda}{|\tau|} \right) \right)}_{f(\beta)}, \end{aligned}$$

where $f(\beta)$ is concave at stationary point since

$$\begin{cases} f'(\beta_*) = 0 \implies 2\lambda\beta_*^2 = 3\lambda \left(\lambda^2 + \frac{9c_\mu}{|\tau|} \right) \\ f''(\beta_*) = \frac{1}{|\beta_*|^3} \left(4\lambda - \frac{12\lambda}{\beta_*^2} \left(\lambda^2 + \frac{9c_\mu}{|\tau|} \right) \right) = \frac{1}{|\beta_*|^3} \left(4\lambda - \frac{12}{3/2} \lambda \right) < 0 \end{cases},$$

then combine with regional condition (I.2), and also apply assumption $c_\lambda \leq \frac{1}{3}$ and $c_\mu \leq \frac{1}{300}$, we gain

$$0.38 \sum_{i \in J} \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i) \geq 0.3 \min_{\beta = \frac{0.79}{\sqrt{|\tau|}}, 0.79} f(\beta)$$

² Apply $\exp[-x^2/2] > 1 - x^2/2$

$$\begin{aligned}
&\geq 0.3 \min \left\{ \frac{2c_\lambda}{0.79} - \frac{c_\lambda^3 + 9c_\mu c_\lambda}{0.79^3}, \lambda \left(\frac{2}{0.79} - \frac{c_\lambda^2 + 9c_\mu}{0.79^3} \right) \right\} \\
&\geq 0.3 \min \{2c_\lambda, 2\lambda\} \geq 0.6\lambda.
\end{aligned} \tag{I.11}$$

3. (Upper bound $\mathbb{E}\chi[\beta]_i$) When $\beta_{(0)}^2 = (|\gamma_{(0)}| - \mu)^2 - \eta$ for some $\eta > 0$. With monotonicity [Lemma F.3](#), which implies:

$$\begin{aligned}
(1 - \mathbb{E}_{\mathbf{s}_{(0)}} \text{erf}_{\beta_{(0)}}(\lambda, \mathbf{s}_{(0)})) &\geq (1 - \mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)})) \\
&\geq (1 - \mathbb{E}_{\mathbf{s}_i} \text{erf}_{\beta_i}(\lambda, \mathbf{s}_i)),
\end{aligned} \tag{I.12}$$

then combine (I.11)-(I.12) and use $\mu \leq \frac{c_\mu}{4\sqrt{|\tau|}}$ from [Lemma E.5](#)

$$\begin{aligned}
\text{(I.10)} &\leq - \underbrace{\left((|\gamma_{(0)}|^2 - \mu)^2 - \beta_{(0)}^2 - \eta \right)}_{=0} (1 - \mathbb{E}_{\mathbf{s}_{(0)}} \text{erf}_{\beta_{(0)}}(\lambda, \mathbf{s}_{(0)})) \\
&\quad + \left(\sum_{i \in \tau \setminus (0)} \beta_i^2 - (|\gamma_{(1)}| - \mu)^2 - \eta \right) \underbrace{(1 - \mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)}))}_{<1} \\
&\quad - 0.38 \sum_{i \in J} \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i) \\
&\leq \left(\|\beta_\tau\|_2^2 - \|\gamma\|_2^2 + 2\mu \|\gamma\|_1 \right) - 0.6\lambda \\
&\leq \frac{2c_\mu}{\sqrt{|\tau|}} - 0.6\lambda.
\end{aligned} \tag{I.13}$$

On the other hand, when $\beta_{(0)}^2 \geq (|\gamma_{(0)}| - \mu)^2 > 0.38$, combining (I.11)-(I.12) gives:

$$\begin{aligned}
\text{(I.10)} &\leq \left(\|\beta_\tau\|_2^2 - \|\gamma\|_2^2 + 2\mu \|\gamma\|_1 \right) + \left((|\gamma_{(0)}| - \mu)^2 - \beta_{(0)}^2 \right) \mathbb{E}_{\mathbf{s}_{(0)}} \text{erf}_{\beta_{(0)}}(\lambda, \mathbf{s}_{(0)}) \\
&\quad + \left((|\gamma_{(1)}| - \mu)^2 - \sum_{i \in \tau \setminus (0)} \beta_i^2 \right) \mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)}) - 0.38 \sum_{i \in J} \mathbb{E}_{\mathbf{s}_i} f_{\beta_i}(\lambda, \mathbf{s}_i) \\
&\leq \left(\frac{c_\mu}{\sqrt{|\tau|}} + 4\mu \right) + \left(\gamma_{(1)}^2 - \|\beta_\tau\|_2^2 + \beta_{(0)}^2 \right) \mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)}) - 0.6\lambda,
\end{aligned} \tag{I.14}$$

where [Lemma F.2](#) provides the upper bound for $\mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)})$ as

$$\begin{aligned}
\mathbb{E}_{\mathbf{s}_{(1)}} \text{erf}_{\beta_{(1)}}(\lambda, \mathbf{s}_{(1)}) &= 1 - \frac{1}{n\theta\beta_{(1)}} \mathbb{E}\chi[\beta]_{(1)} \leq 1 - \frac{\sigma_{(1)}}{n\theta|\beta_{(1)}|} \mathbb{E}\chi[\beta]_{(1)} \\
&= 1 - \frac{1}{|\beta_{(1)}|} \left(|\beta_{(1)}| - \sqrt{\frac{2}{\pi}}\lambda \right) \leq \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{|\beta_{(1)}|},
\end{aligned} \tag{I.15}$$

then calculate the constant for the second term in (I.14) by writing $\kappa = \left| \frac{\gamma_{(1)}}{\gamma_{(0)}} \right| = \left| \frac{\beta_{(0)}}{\beta_{(1)}} \right| \leq \frac{5}{4}$, which provides $\gamma_{(1)}^2 \leq \frac{(1+\mu)\kappa^2}{\kappa^2+1}$ and $\beta_{(0)}^2 \leq \frac{\|\beta_\tau\|_2^2 \kappa^2}{\kappa^2+1}$ where $\mu < \frac{c_\mu}{4}$, and by applying $|\beta_{(1)}| > \frac{4}{5} |\beta_{(0)}| \geq 0.3$, we have

$$\frac{(\gamma_{(1)}^2 - 1) + c_\mu + \beta_{(0)}^2}{|\beta_{(1)}|} \leq -\frac{\kappa}{(\kappa^2 + 1)|\beta_{(0)}|} + \kappa |\beta_{(0)}| + \frac{\mu + c_\mu}{0.3}$$

$$\leq \frac{\kappa^2 - 1}{\sqrt{\kappa^2 + 1}} + \kappa \left(\|\boldsymbol{\beta}_\tau\|_2^2 - 1 \right) + 4.2c_\mu \leq 0.36 + 6c_\mu, \quad (\text{I.16})$$

and finally combine (I.15)-(I.16), follow from (I.14) and use $c_\lambda \leq \frac{1}{3}$:

$$\begin{aligned} (\text{I.10}) &\leq \frac{2c_\mu}{\sqrt{|\boldsymbol{\tau}|}} + \sqrt{\frac{2}{\pi}} \left(\gamma_{(1)}^2 - 1 + c_\mu + \beta_{(0)}^2 \right) \frac{\lambda}{|\boldsymbol{\beta}_{(1)}|} - 0.6\lambda \\ &\leq \frac{2c_\mu}{\sqrt{|\boldsymbol{\tau}|}} + \sqrt{\frac{2}{\pi}} \left(0.36\lambda + \frac{6c_\mu c_\lambda}{0.3} \right) - 0.6\lambda \\ &\leq \frac{4c_\mu}{\sqrt{|\boldsymbol{\tau}|}} - 0.3\lambda \end{aligned} \quad (\text{I.17})$$

3. (Collect all results) Combine the components of pseudo Hessian (I.7), (I.9) with bounds for $g_1 + g_2$ from (I.13) and (I.17), and use Lemma E.5 which provides both $\mu p \theta |\boldsymbol{\tau}| < \frac{c_\mu}{4}$ and $\theta |\boldsymbol{\tau}| < \frac{c_\mu}{4}$ where $c_\mu < \frac{1}{300}$ and $c_\lambda \geq \frac{1}{5}$, we can obtain:

$$\begin{aligned} \mathbf{v}^* \widetilde{\text{Hess}}_{\varphi_{\rho_1}}[\mathbf{a}]\mathbf{v} &\leq g_1(\boldsymbol{\beta}) + g_2(\boldsymbol{\beta}) + \frac{7c_\mu n \theta}{\sqrt{|\boldsymbol{\tau}|}} + (18\mu p + 8) n \theta^2 \\ &\leq n \theta \cdot \left(\frac{4c_\mu}{\sqrt{|\boldsymbol{\tau}|}} - 0.3\lambda \right) + n \theta \cdot \frac{7c_\mu}{\sqrt{|\boldsymbol{\tau}|}} + n \theta \cdot \frac{6.5c_\mu}{|\boldsymbol{\tau}|} \\ &\leq \frac{n \theta}{\sqrt{|\boldsymbol{\tau}|}} (0.059 - 0.06) \leq -0.001 n \theta \lambda \end{aligned} \quad (\text{I.18})$$

Finally, the curvature is negative along \mathbf{v} direction with probability at least

$$1 - \underbrace{\mathbb{P}[\mathcal{E}_{\text{cross}}^c]}_{\text{Lemma D.4}} - \underbrace{\mathbb{P}[\mathcal{E}_{\text{ncurv}}^c]}_{\text{Corollary G.3}} - \underbrace{\mathbb{P}[\mathcal{E}_{\chi}^c]}_{\text{Corollary F.4}}. \quad (\text{I.19})$$

Similarly for objective φ_ρ , we have that

Corollary I.2 (Negative curvature for φ_ρ). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda / \sqrt{k}$ in φ_ρ where $c_\lambda \in [\frac{1}{5}, \frac{1}{4}]$, then there exists some numerical constants $C, c, c', c'', \bar{c} > 0$ such that if ρ is δ -smoothed ℓ^1 function where $\delta \leq c'' \lambda \theta^8 / p^2 \log^2 n$, $n > Cp^5 \theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\boldsymbol{\tau}| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ satisfying $|\boldsymbol{\beta}_{(1)}| \geq \frac{4}{5} |\boldsymbol{\beta}_{(0)}|$: for $\mathbf{v} \in \mathcal{S}_{\{(0), (1)\}} \cap \mathbb{S}^{p-1} \cap \mathbf{a}^\perp$,*

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_\rho](\mathbf{a})\mathbf{v} \leq -cn\theta\lambda \quad (\text{I.20})$$

Proof. Choose $\mathbf{v} \in \mathbb{S}^{p-1}$ according to Lemma I.1 and (H.23) from Lemma H.6 with constant multiplier δ satisfies $c''^{1/4} < 10^{-3}c$, we gain

$$\mathbf{v}^* \text{Hess}[\varphi_\rho](\mathbf{a})\mathbf{v} \leq -cn\theta\lambda + 200c'n\theta^2 \leq -cn\theta\lambda/2 \quad (\text{I.21})$$

I.2 Large gradient

For any $\mathbf{a} \in \mathbb{S}^{p-1}$ near subspace and the second largest correlation $\beta_{(1)}$ much smaller than the first correlation $\beta_{(0)}$ while not being near 0, the negative gradient of $\varphi_\rho(\mathbf{a})$ will point at the largest shift. We show this in Lemma I.3, and the φ_ρ version in Corollary I.4 when ρ is properly defined as in Appendix H.

Lemma I.3 (Large gradient for φ_{ℓ^1}). Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_{ℓ^1} with some $c_\lambda \in [\frac{1}{5}, \frac{1}{4}]$, then there exists some numerical constants $C, c', c, \bar{c} > 0$, such that if $n > Cp^5\theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ satisfying $\frac{4}{5} |\boldsymbol{\beta}_{(0)}| > |\boldsymbol{\beta}_{(1)}| > \frac{1}{4 \log \theta^{-1}} \lambda$,

$$\langle \boldsymbol{\sigma}_{(0)} \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0], -\text{grad}[\varphi_{\ell^1}](\mathbf{a}) \rangle \geq cn\theta (\log^{-2} \theta^{-1}) \lambda^2 \quad (\text{I.22})$$

where $\boldsymbol{\sigma}_i = \text{sign}(\boldsymbol{\beta}_i)$.

Proof. 1. (Properties for $\boldsymbol{\alpha}, \boldsymbol{\beta}$) Define $\theta_{\log} = \frac{1}{\log \theta^{-1}}$, we first derive upper bound on the dominant entry $|\boldsymbol{\beta}_{(0)}|$ as follows. Write the geodesic distance between \mathbf{a} and $\boldsymbol{\iota}^* s_i[\mathbf{a}_0]$ as a function of $\boldsymbol{\beta}_i$ as $d_{\mathbb{S}}(\mathbf{a}, \pm \boldsymbol{\iota}^* s_i[\mathbf{a}_0]) = \cos^{-1}(\boldsymbol{\beta}_i)$, then by triangle inequality we have:

$$\begin{aligned} d_{\mathbb{S}}(\mathbf{a}, \pm \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0]) &\geq d_{\mathbb{S}}(\pm \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0], \boldsymbol{\iota}^* s_{(1)}[\mathbf{a}_0]) - d_{\mathbb{S}}(\mathbf{a}, \boldsymbol{\iota}^* s_{(1)}[\mathbf{a}_0]) \\ \implies \cos^{-1} \pm \boldsymbol{\beta}_{(0)} &\geq \cos^{-1} \mu - \cos^{-1} |\boldsymbol{\beta}_{(1)}| \\ \implies \pm \boldsymbol{\beta}_{(0)} &\leq \cos(\cos^{-1} \mu - \cos^{-1} |\boldsymbol{\beta}_{(1)}|) = \mu |\boldsymbol{\beta}_{(1)}| + \sqrt{(1 - \mu^2)(1 - \boldsymbol{\beta}_{(1)}^2)} \\ &\leq 1 - \frac{1}{2} (|\boldsymbol{\beta}_{(1)}| - \mu)^2. \end{aligned}$$

Use the regional condition $|\boldsymbol{\beta}_{(1)}| \geq \frac{\theta_{\log}}{4} \lambda$ and since $\mu |\tau|^{3/2} < \frac{c_\lambda}{100} \theta_{\log}$ from [Definition E.1](#), implies

$$|\boldsymbol{\beta}_{(0)}| \leq 1 - \frac{\boldsymbol{\beta}_{(1)}^2}{2} \left(1 - \frac{4\mu\sqrt{|\tau|}}{\theta_{\log} c_\lambda}\right) \leq 1 - 0.49 \boldsymbol{\beta}_{(1)}^2 =: \beta_{\text{ub}}. \quad (\text{I.23})$$

Meanwhile a lower bound for $\boldsymbol{\beta}_{(0)}$ can be easily determined by the other side of regional condition:

$$|\boldsymbol{\beta}_{(0)}| \geq \frac{5}{4} |\boldsymbol{\beta}_{(1)}| =: \beta_{\text{lb}}. \quad (\text{I.24})$$

Also since $\boldsymbol{\beta} = \mathbf{M}\boldsymbol{\alpha}$, based on properties of \mathbf{M} from [Lemma E.4](#). When $\|\boldsymbol{\alpha}_\tau\|_2 \leq 1 + c_\mu$ and $\|\boldsymbol{\alpha}_{\tau^c}\|_2 \leq \gamma \leq \frac{c_\mu \theta_{\log}^2}{4\mu\sqrt{p}|\tau|}$, we gain:

$$\begin{aligned} \boldsymbol{\beta}_{(0)} &= \boldsymbol{\alpha}_{(0)} + \mathbf{e}_{(0)}^* \mathbf{M} \boldsymbol{\alpha}_{\setminus(0)} \\ \implies |\boldsymbol{\alpha}_{(0)} - \boldsymbol{\beta}_{(0)}| &\leq \mu \sqrt{|\tau|} \|\boldsymbol{\alpha}_\tau\|_2 + \mu \sqrt{p} \|\boldsymbol{\alpha}_{\tau^c}\|_2 \\ &\leq \frac{c_\mu \theta_{\log}^2 (1 + c_\mu)}{4|\tau|} + \mu \sqrt{p} \gamma \leq \frac{c_\mu \theta_{\log}^2}{|\tau|}. \end{aligned} \quad (\text{I.25})$$

and therefore $|\boldsymbol{\alpha}_{(0)}| \leq |\boldsymbol{\beta}_{(0)}| + \frac{c_\mu \theta_{\log}^2}{|\tau|} \leq 1 - .49 \left(\frac{\theta_{\log}}{4} \lambda\right)^2 + \frac{c_\mu \theta_{\log}^2}{|\tau|} < 1$.

2. (Upper bound of $\boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}]$) Define a piecewise smooth convex upper bound h for $\boldsymbol{\beta}_i \boldsymbol{\chi}[\boldsymbol{\beta}]_i$ as:

$$h(\boldsymbol{\beta}_i) := \begin{cases} \boldsymbol{\beta}_i^2 - \frac{\nu_1 \lambda}{2} |\boldsymbol{\beta}_i| & |\boldsymbol{\beta}_i| \geq \nu_1 \lambda \\ \frac{1}{2} \boldsymbol{\beta}_i^2 & |\boldsymbol{\beta}_i| \leq \nu_1 \lambda \end{cases},$$

then [Lemma N.7](#) tells us since $\|\boldsymbol{\beta}_{\tau \setminus (0)}\|_\infty \leq \boldsymbol{\beta}_{(1)}$:

$$\begin{aligned} \sum_{i \in \tau \setminus (0)} h(\boldsymbol{\beta}_i) &\leq \|\boldsymbol{\beta}_{\tau \setminus (0)}\|_2^2 \left(1 - \frac{\nu_1 \lambda \boldsymbol{\beta}_{(1)}}{2 \boldsymbol{\beta}_{(1)}^2}\right) \leq \left(1 + \frac{c_\mu \theta_{\log}^2}{|\tau|} - \boldsymbol{\beta}_{(0)}^2\right) \left(1 - \frac{\nu_1 \lambda}{2 \boldsymbol{\beta}_{(1)}}\right) \\ &\leq \left(1 - \frac{\nu_1 \lambda}{2 \boldsymbol{\beta}_{(1)}}\right) \left(1 - \boldsymbol{\beta}_{(0)}^2\right) + \frac{c_\mu \theta_{\log}^2}{|\tau|}, \end{aligned}$$

then condition on the following event using [Corollary F.4](#),

$$\mathcal{E}_{\bar{\chi}} := \left\{ \beta_i \chi[\beta]_i \leq \begin{cases} n\theta \cdot h(\beta_i) + \frac{c_\mu \theta}{p^{3/2}} |\beta_i|, & \forall i \in \tau \setminus (0) \\ n\theta \cdot 4\beta_i^2 \theta |\tau| + \frac{c_\mu \theta}{p^{3/2}} |\beta_i|, & \forall i \in \tau^c \end{cases} \right\},$$

which provides the upper bound of $\beta^* \chi[\beta]$ by applying $5p > \log^{8/3}(p \log^2 p) > (\theta_{\log}^2)^{4/3}$ from lower bound of θ from [Definition E.1](#), $\|\beta_{\tau^c}\|_2 \leq \frac{c_\mu \theta_{\log}}{\sqrt{\theta} |\tau|}$ from [Lemma E.5](#), $|\tau| \leq \sqrt{p}$ from lemma assumption and let $c_\mu < \frac{1}{100}$:

$$\begin{aligned} \beta^* \chi[\beta] &\leq \chi[\beta]_{(0)} \beta_{(0)} + \sum_{i \in \tau \setminus (0)} \beta_i \chi[\beta]_i + \langle \beta_{\tau^c}, \chi[\beta]_{\tau^c} \rangle \\ &\leq \chi[\beta]_{(0)} \beta_{(0)} + n \left(\theta \sum_{i \in \tau \setminus (0)} h(\beta_i) + 4\theta^2 |\tau| \|\beta_{\tau^c}\|_2^2 \right. \\ &\quad \left. + \frac{c_\mu \theta}{p^{3/2}} \left(\sqrt{|\tau|} \|\beta_\tau\|_2 + \sqrt{p} \|\beta_{\tau^c}\|_2 \right) \right) \\ &\leq \chi[\beta]_{(0)} \beta_{(0)} + n \left(\theta \cdot \eta (1 - \beta_{(0)}^2) + \theta \cdot \frac{c_\mu \theta_{\log}^2}{|\tau|} + \frac{4\theta^2 |\tau| c_\mu^2 \theta_{\log}^2}{\theta |\tau|^2} \right. \\ &\quad \left. + c_\mu \theta \left(\frac{1 + c_\mu}{p^{3/4} |\tau|} + \frac{c_\mu \theta_{\log}}{p \sqrt{\theta} |\tau|} \right) \right) \\ &\leq \chi[\beta]_{(0)} \beta_{(0)} + n\theta \left(\eta (1 - \beta_{(0)}^2) + \frac{6c_\mu \theta_{\log}^2}{|\tau|} \right), \end{aligned} \tag{I.26}$$

where $\eta = 1 - \frac{\nu_1 \lambda}{2\beta_{(1)}}$.

3. (Align the gradient with $\iota^* s_{(0)}[a_0]$) Base on the definition β , since $\beta_{(0)} = \langle a, \iota^* s_{(0)}[a_0] \rangle$, we can expect that the negative gradient is likely aligned with direction toward one of the candidate solution $\pm \iota^* s_{(0)}[a_0]$. Wlog assume that both $\beta_{(0)}, \beta_{(1)}$ are positive, then expand the gradient and use incoherent property for a_0 [Lemma E.4](#) we have:

$$\begin{aligned} \left\langle \iota^* s_{(0)}[a_0], -\text{grad}_{\varphi_{\epsilon_1}}[a] \right\rangle &= \langle \iota^* s_{(0)}[a_0], \iota^* C_{a_0} (\chi[\beta] - \beta^* \chi[\beta] \alpha) \rangle \\ &\geq (\chi[\beta]_{(0)} - \beta^* \chi[\beta] \alpha_{(0)}) - \mu \|\chi[\beta]_{\setminus (0)} - \beta^* \chi[\beta] \alpha_{\setminus (0)}\|_1, \end{aligned} \tag{I.27}$$

where $\setminus (0)$ is an abbreviation of the complement set $[\pm 2p_0] \setminus (0)$. The latter part of (I.27) has an upper bound using bounds of $\beta^* \chi[\beta] < \frac{3n\theta}{2}$, $\|\chi[\beta]_{\tau^c}\|_2 < \frac{n\theta\gamma_2}{20}$ from (I.62), and $\|\chi[\beta]_{\tau \setminus (0)}\|_2 \leq n\theta \|\beta_{\tau \setminus (0)}\|_2$ in event $\mathcal{E}_{\bar{\chi}}$, we obtain:

$$\begin{aligned} &\mu \|\chi[\beta]_{\setminus (0)} - \beta^* \chi[\beta] \alpha_{\setminus (0)}\|_1 \\ &\leq \mu \left(\sqrt{|\tau|} \|\chi[\beta]_{\tau \setminus (0)}\|_2 + \beta^* \chi[\beta] \sqrt{|\tau|} \|\alpha_{\tau \setminus (0)}\|_2 \right. \\ &\quad \left. + \sqrt{p} \|\chi[\beta]_{\tau^c}\|_2 + \beta^* \chi[\beta] \sqrt{p} \|\alpha_{\tau^c}\|_2 \text{ Big} \right) \\ &\leq n\theta \cdot \left[\mu \sqrt{|\tau|} (\|\beta_\tau\|_2 - |\beta_{(0)}|) + \mu \sqrt{|\tau|} (\|\alpha_\tau\|_2 - |\alpha_{(0)}|) \right. \\ &\quad \left. + \frac{1}{20} \mu \sqrt{p} \gamma_2 + \frac{3}{2} \mu \sqrt{p} \gamma_2 \right] \\ &\leq n\theta \cdot \frac{c_\mu \theta_{\log}^2}{4|\tau|} \left[2(1 + c_\mu) - |\beta_{(0)}| - |\alpha_{(0)}| + \left(\frac{1}{20} + \frac{3}{2} \right) c_\mu \right] \\ &\leq n\theta \cdot \frac{c_\mu \theta_{\log}^2}{|\tau|} (0.5 + c_\mu - 0.5\beta_{(0)}). \end{aligned} \tag{I.28}$$

On the other hand, the former term of (I.27) possesses a lower bound using (I.25)-(I.26), $\chi[\beta]_{(0)} > n\theta \left(\beta_{(0)} - \frac{\nu_1}{2}\lambda - \frac{c_\mu}{p} \right) \geq n\theta (\beta_{(0)} - 0.51\nu_1\lambda)$ and $\alpha_{(0)} \leq 1$:

$$\begin{aligned}
& \chi[\beta]_{(0)} - \beta^* \chi[\beta] \alpha_{(0)} \\
& \geq (1 - \alpha_{(0)}\beta_{(0)}) \chi[\beta]_{(0)} - n\theta \cdot \left[\eta (1 - \beta_{(0)}^2) + \frac{6c_\mu\theta_{\log}^2}{|\tau|} \right] \alpha_{(0)} \\
& \geq n\theta \underbrace{\left(1 - \left(\beta_{(0)} + \frac{c_\mu\theta_{\log}^2}{|\tau|} \right) \beta_{(0)} \right)}_{(a)} (\beta_{(0)} - 0.51\nu_1\lambda) \\
& \quad - n\theta \underbrace{\left[\eta (1 - \beta_{(0)}^2) \left(\beta_{(0)} + \frac{c_\mu\theta_{\log}^2}{|\tau|} \right) + \frac{6c_\mu\theta_{\log}^2}{|\tau|} \alpha_{(0)} \right]}_{(b)} \\
& \geq n\theta \underbrace{\left[(1 - \beta_{(0)}^2) (\beta_{(0)} - 0.51\nu_1\lambda) - \frac{c_\mu\theta_{\log}^2\beta_{(0)}^2}{|\tau|} \right]}_{(a)} \\
& \quad - \underbrace{\left(1 - \beta_{(0)}^2 \right) \eta \beta_{(0)} - \eta \frac{c_\mu\theta_{\log} (1 - \beta_{(0)}^2)}{|\tau|} - \frac{6c_\mu\theta_{\log}^2}{|\tau|}}_{(b)} \\
& \geq n\theta \left[(1 - \beta_{(0)}^2) ((1 - \eta) \beta_{(0)} - 0.51\nu_1\lambda) - \frac{c_\mu\theta_{\log}^2}{|\tau|} ((1 - \eta)\beta_{(0)}^2 + 7) \right], \tag{I.29}
\end{aligned}$$

combine (I.27) with (I.28)-(I.29) and $\eta > 0$, we have

$$\begin{aligned}
\text{(I.27)} & \geq n\theta \left[(1 - \beta_{(0)}^2) ((1 - \eta) \beta_{(0)} - 0.51\nu_1\lambda) - \frac{c_\mu\theta_{\log}^2}{|\tau|} ((1 - \eta)\beta_{(0)}^2 + 7) \right] \\
& \quad - n\theta \cdot \frac{c_\mu\theta_{\log}^2}{|\tau|} (0.5 + c_\mu - 0.5\beta_{(0)}) \\
& \geq n\theta \left[\underbrace{\left(1 - \beta_{(0)}^2 \right) \left(\frac{\nu_1\lambda}{2\beta_{(1)}} \beta_{(0)} - 0.51\nu_1\lambda \right)}_{f(\beta)} - \frac{8c_\mu\theta_{\log}^2}{|\tau|} \text{Big} \right]. \tag{I.30}
\end{aligned}$$

4. (Lower bound of $f(\beta)$) Given a fixed $\beta_{(1)}$, the cubic function $f(\beta_{(0)})$ has zeros set $\beta_{(0)} \in \{\pm 1, 1.02\beta_{(1)}\}$ and has negative leading coefficient. Combine with the condition of $\beta_{(0)} \in \{\beta_{\text{lb}}, \beta_{\text{ub}}\}$ from (I.23)-(I.24), we can observe that

$$\beta_{(0)} \in [\beta_{\text{lb}}, \beta_{\text{ub}}] = \left[\frac{5}{4}\beta_{(1)}, 1 - 0.49\beta_{(1)}^2 \right] \subseteq [1.02\beta_{(1)}, 1],$$

therefore the cubic term is always positive and minimizer is either one of the boundary point. When $\beta_{(0)} = \beta_{\text{lb}}$, use $(1 + \frac{25}{16})\beta_{(1)}^2 < 1.01$, and use $\nu_1\lambda < \frac{\sqrt{\theta_{\log}}}{2\sqrt{|\tau|}} \leq \frac{1}{2\sqrt{2}}$, since $|\tau| \geq 2$, we have:

$$f(\beta_{\text{lb}}) \geq (1 - \beta_{\text{lb}}^2) \left(\frac{\nu_1\lambda}{2\beta_{(1)}} \beta_{\text{lb}} - 0.51\nu_1\lambda \right) \geq (1 - 0.616) \cdot \left(\frac{5}{8} - 0.51 \right) \nu_1\lambda$$

$$\geq \frac{1}{16\sqrt{2}}\nu_1\lambda \geq \frac{\theta_{\log}^2}{32}\lambda^2. \quad (\text{I.31})$$

On the other hand when $\beta_{(0)} = \beta_{\text{ub}}$:

$$\begin{aligned} f(\beta_{\text{ub}}) &\geq (1 - \beta_{\text{ub}}^2) \left(\frac{\nu_1\lambda}{2\beta_{(1)}}\beta_{\text{ub}} - 0.51\nu_1\lambda \right) \\ &\geq 0.49\beta_{(1)}^2 \cdot \left(\frac{\nu_1\lambda}{2\beta_{(1)}} \left(1 - 0.49\beta_{(1)}^2 \right) - 0.51\nu_1\lambda \right), \end{aligned}$$

which is a cubic function of $\beta_{(1)}$ with negative leading coefficient, whose zeros set is $\{-0.73, 0, 2.81\}$. Thus it minimizes at the boundary points of $\beta_{(1)} \in \left[\frac{\lambda}{4\log\theta^{-1}}, 1 \right] \subset [0, 2.81]$, thus assign $\beta_{(1)} = \frac{\lambda}{4\log\theta^{-1}}$, we have:

$$\begin{aligned} f(\beta_{\text{ub}}) &\geq 0.49 \left(\frac{\lambda}{4\log\theta^{-1}} \right)^2 \cdot \left(\frac{1}{2} \left(1 - 0.49 \left(\frac{\lambda}{4\log\theta^{-1}} \right)^2 \right) - 0.51\nu_1\lambda \right) \\ &\geq \frac{1}{6} \left(\frac{\lambda}{4\log\theta^{-1}} \right)^2 \geq \frac{\theta_{\log}^2}{96}\lambda^2. \end{aligned} \quad (\text{I.32})$$

Finally combine (I.30) with the lower bound of cubic function (I.31)-(I.32) together with condition $c_\mu < \frac{c_\lambda^2}{800}$ and $\nu_1 = \frac{\sqrt{\theta_{\log}}}{2}$, obtain

$$\begin{aligned} \langle \iota^* s_{(0)}[\mathbf{a}_0], -\text{grad}_{\varphi_{\ell_1}}[\mathbf{a}] \rangle &\geq n\theta \cdot \left(\min\{f(\beta_{\text{ub}}), f(\beta_{\text{lb}})\} - \frac{8c_\mu\theta_{\log}^2}{|\boldsymbol{\tau}|} \right) \\ &\geq n\theta \left(\frac{\theta_{\log}^2 c_\lambda^2}{96|\boldsymbol{\tau}|} - \frac{8\theta_{\log}^2 c_\lambda^2}{800|\boldsymbol{\tau}|} \right) \geq 6 \times 10^{-3} n\theta\theta_{\log}^2 c_\lambda^2. \end{aligned} \quad (\text{I.33})$$

The proof for the case where $\beta_{(0)}$ negative can be derived in the same manner. ■

As a consequence, we have that

Corollary I.4 (Large gradient for φ_ρ). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_ρ with $c_\lambda \in [\frac{1}{5}, \frac{1}{4}]$, then there exists some numerical constants $C, c, c', c'', \bar{c} > 0$ such that if ρ is δ -smoothed ℓ^1 function where $\delta \leq c'\lambda\theta^8/p^2\log^2 n$ with $n > Cp^5\theta^{-2}\log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\boldsymbol{\tau}| \leq k} \mathfrak{A}(\mathcal{S}_\boldsymbol{\tau}, \gamma(c_\mu))$ satisfying $\frac{4}{5}|\beta_{(0)}| > |\beta_{(1)}| > \frac{1}{4\log\theta^{-1}\lambda}$,*

$$\langle \boldsymbol{\sigma}_{(0)} \iota^* s_{(0)}[\mathbf{a}_0], -\text{grad}[\varphi_\rho](\mathbf{a}) \rangle \geq cn\theta (\log^{-2}\theta^{-1}) \lambda^2 \quad (\text{I.34})$$

where $\sigma_i = \text{sign}(\beta_i)$.

Proof. Choose $\iota^* s_{(0)}[\mathbf{a}_0]$ as in Lemma I.3, and apply (H.22) from Lemma H.6 with the constant multiplier of δ satisfies $c''^4 < c/4$, then utilize $\theta|\boldsymbol{\tau}|\log^2\theta^{-1} < c_\mu$ from Definition E.1 we have

$$\langle \boldsymbol{\sigma}_{(0)} \iota^* s_{(0)}[\mathbf{a}_0], -\text{grad}[\varphi_\rho](\mathbf{a}) \rangle \geq cn\theta(\log^{-2}\theta^{-1})\lambda - c''n\theta^2 \geq cn\theta(\log^{-2}\theta^{-1})\lambda/2 \quad (\text{I.35})$$

I.3 Convex near solutions

For any $\mathbf{a} \in \mathbb{S}^{p-1}$ near subspace and the second largest correlation $\beta_{(1)}$ smaller than $\frac{1}{4\log\theta^{-1}}\lambda$, then φ_ρ will be strongly convex at \mathbf{a} . We show this in Lemma I.5, and the φ_ρ version in Corollary I.6 when ρ is properly defined as in Appendix H.

Lemma I.5 (Strong convexity of φ_{ℓ^1} near shift). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_{ℓ^1} with $c_\lambda \in [\frac{1}{4}, \frac{1}{5}]$, then there exists some numerical constants $C, c, c'\bar{c} > 0$ such that if $n > Cp^5\theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ satisfying $|\beta_{(1)}| < \frac{1}{4 \log \theta^{-1}} \lambda$: for all $\mathbf{v} \in \mathbb{S}^{p-1} \cap \mathbf{v}^\perp$,*

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) \mathbf{v} > cn\theta; \quad (\text{I.36})$$

furthermore, there exists $\bar{\mathbf{a}}$ as an local minimizer such that

$$\min_{\ell} \|\bar{\mathbf{a}} - s_\ell[\mathbf{a}_0]\|_2 \leq \frac{1}{2} \max\{\mu, p^{-1}\}. \quad (\text{I.37})$$

Proof. 1. (Expectation of χ near shifts) We will write \mathbf{x} as \mathbf{x}_0 through out this proof. When \mathbf{a} is near one of the shift, the χ operator shrinks all other smaller entries of correlation vector $\beta_{\setminus(0)}$ in an even larger shrinking ratio. Firstly we can show $|\langle \beta_{\setminus(0)}, \mathbf{x}_{\setminus(0)} \rangle|$ is no larger then $\lambda/2$ with probability at least $1 - 4\theta$, since

$$\begin{aligned} & \mathbb{P} \left[|\langle \beta_{\setminus(0)}, \mathbf{x}_{\setminus(0)} \rangle| > \frac{\lambda}{2} \right] \\ & \leq \mathbb{P} \left[|\langle \beta_{\tau \setminus(0)}, \mathbf{x}_{\tau \setminus(0)} \rangle| > \frac{2\lambda}{5} \right] + \mathbb{P} \left[|\langle \beta_{\tau^c}, \mathbf{x}_{\tau^c} \rangle| > \frac{\lambda}{10} \right] \leq 4\theta \end{aligned} \quad (\text{I.38})$$

via [Corollary E.6](#) and [Corollary E.7](#). Now recall from [Lemma F.2](#) and the derivation of [\(F.10\)](#)-[\(F.11\)](#), we know for every $i \neq (0)$,

$$\begin{aligned} \sigma_i \mathbb{E} \chi[\beta]_i &= n\theta |\beta_i| \mathbb{E}_{\mathbf{s}_i} [1 - \text{erf}_{\beta_i}(\lambda, \mathbf{s}_i)] \\ &\leq n\theta |\beta_i| \mathbb{E}_{g, \mathbf{x}_{\setminus i}} \left[g^2 \mathbf{1}_{\{|\beta_i g + \beta_{(0)} \mathbf{x}_{(0)} + \beta_{\setminus\{(0), i\}}^* \mathbf{x}_{\setminus\{(0), i\}}| > \lambda\}} \right] \\ &\leq n\theta |\beta_i| \left(\mathbb{E} g^2 \mathbf{1}_{\{|\beta_i g| > \frac{\lambda}{2}\}} + \mathbb{P}[\mathbf{x}_{(0)} \neq 0] \right. \\ &\quad \left. + \mathbb{P} \left[|\langle \beta_{\setminus\{(0), i\}}, \mathbf{x}_{\setminus\{(0), i\}} \rangle| > \frac{\lambda}{2} \right] \right) \\ &\leq n\theta |\beta_i| \left((\mathbb{E} g^2)^{1/2} \mathbb{P} \left[|\beta_{(1)} g| > \frac{\lambda}{2} \right]^{1/2} + \theta + 4\theta \right) \\ &\leq n\theta |\beta_i| (\exp(-\log^2 \theta^{-1}) + 5\theta) \\ &\leq 6n\theta^2 |\beta_i| \end{aligned} \quad (\text{I.39})$$

where the third inequality is derived using union bound; the the fourth inequality is the result of [\(I.38\)](#), and the fifth inequality is derived from Gaussian tail bound [lemma N.1](#).

2. (Local strong convexity) Let $\gamma = C_{\mathbf{a}_0}^* \boldsymbol{\iota} \mathbf{v}$, for any $\|\mathbf{v}\|_2 = 1$ we have $\|\gamma\|_2^2 \leq 1 + \mu p$. Furthermore:

$$\begin{aligned} |\gamma_{(0)}| &= |\langle \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0], \mathbf{v} \rangle| = |\langle \mathbf{P}_{\mathbf{a}_0} \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0], \mathbf{v} \rangle| = |\langle \boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0] - \beta_{(0)} \mathbf{a}, \mathbf{v} \rangle| \\ &\leq \|\boldsymbol{\iota}^* s_{(0)}[\mathbf{a}_0] - \beta_{(0)} \mathbf{a}\|_2 \leq \sqrt{1 - \beta_{(0)}^2}. \end{aligned} \quad (\text{I.40})$$

Consider any such \mathbf{v} , the pseudo Hessian can be lower bounded as

$$\begin{aligned} \mathbf{v}^* \widetilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v} &= -\gamma^* \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \gamma \\ &\geq -\gamma_{(0)}^2 \left\| \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_{(0)} \right\|_2^2 - \sum_{i \neq (0)} \left\| \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_i \right\|_2^2 \gamma_i^2 \\ &\quad - 2 \sum_{i \neq j} \left| \mathbf{e}_i^* \check{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \check{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| |\gamma_i| |\gamma_j| \end{aligned}$$

$$\begin{aligned}
&\geq -\left(1 - \beta_{(0)}^2\right) \|\mathbf{x}\|_2^2 - \max_{i \neq (0)} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}]\|_2^2 \|\boldsymbol{\gamma}\|_2^2 \\
&\quad - 2 \max_{i \neq j} \left| \mathbf{e}_i^* \tilde{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \tilde{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| \|\boldsymbol{\gamma}\|_1^2,
\end{aligned} \tag{I.41}$$

where the second term is bounded by using its expectation derived in [Lemma G.2](#), and utilize $\mathbb{P}[|s_i| > \lambda/2] < 4\theta$ from [\(I.38\)](#), $\mathbb{E}\boldsymbol{\chi}$ from [\(I.39\)](#) and regional condition $|\beta_{(1)}| \leq \frac{\lambda}{4 \log \theta^{-1}}$ to acquire

$$\begin{aligned}
\mathbb{E} \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}]\|_2^2 &= n\theta [1 - \mathbb{E}_{s_i} \operatorname{erf} \beta_i(\lambda, s_i) + \mathbb{E}_{s_i} f_{\beta_i}(\lambda, s_i)] \\
&\leq \frac{|\mathbb{E}\boldsymbol{\chi}[\beta]_i|}{|\beta_i|} + n\theta \cdot \left(\max_{|s_i| \leq \frac{\lambda}{2}} f_{\beta_i}(\lambda, s_i) + \mathbb{P}\left[|s_i| > \frac{\lambda}{2}\right] \right) \\
&\leq 6n\theta^2 + \frac{2n\theta}{\sqrt{2\pi}} \max_{|s_i| \leq \frac{\lambda}{2}} \left(\frac{\lambda + |s_i|}{|\beta_i|} \cdot \exp\left[-\frac{(\lambda - |s_i|)^2}{2\beta_i^2}\right] \right) + 4n\theta^2 \\
&\leq 10n\theta^2 + n\theta \cdot \log \theta^{-1} \exp(-2 \log^2 \theta^{-1}) \\
&\leq 11n\theta^2,
\end{aligned} \tag{I.42}$$

and define the events $\mathcal{E}_{\|\mathbf{x}\|_2}$, $\mathcal{E}_{\text{cross}}$ and $\mathcal{E}_{\text{pcurv}}$ as follows:

$$\begin{cases} \mathcal{E}_{\text{pcurv}} := \left\{ \forall \mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_{\tau}, \gamma(c_{\mu})), \|\mathbf{P}_{I(\mathbf{a})} s_{-i}[\mathbf{x}]\|_2^2 \leq 11n\theta^2 + \frac{c_{\mu}n\theta}{p} \right\} \\ \mathcal{E}_{\text{cross}} := \left\{ \forall \mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_{\tau}, \gamma(c_{\mu})), |\beta_{(1)}| \leq \frac{\lambda}{4 \log \theta^{-1}}, \max_{i \neq j \in [\pm p]} \left| \mathbf{e}_i^* \tilde{\mathbf{C}}_{\mathbf{x}} \mathbf{P}_{I(\mathbf{a})} \tilde{\mathbf{C}}_{\mathbf{x}} \mathbf{e}_j \right| \leq 8n\theta^3 \right\} \\ \mathcal{E}_{\|\mathbf{x}\|_2} := \left\{ \|\mathbf{x}\|_2^2 \leq n\theta + 3\sqrt{n\theta} \log n \right\} \end{cases} \quad . \tag{I.43}$$

For the Hessian term, on the event $\mathcal{E}_{\text{pcurv}} \cap \mathcal{E}_{\text{cross}} \cap \mathcal{E}_{\|\mathbf{x}\|_2}$, and use all $\mu p^2 \theta^2$, $\mu p \theta |\tau|$ and $\theta \sqrt{p}$ are all less than $\frac{c_{\mu}}{4 \log^2 \theta^{-1}}$, from [Lemma E.5](#), and from lemma assumption with sufficiently large C we have $n > \theta^{-1} 36 \log^2 n$, thus $\mathbf{v}^* \tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v}$ can be lower bounded from [\(I.41\)](#) as

$$\begin{aligned}
\mathbf{v}^* \tilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v} &\geq -\left(1 - \beta_{(0)}^2\right) \left(n\theta + 3\sqrt{n\theta} \log n\right) \\
&\quad - (1 + \mu p) \left(11n\theta^2 + \frac{c_{\mu}n\theta}{p}\right) - 8p(1 + \mu p) \cdot 8n\theta^3 \\
&\geq -\frac{1}{2}n\theta \cdot (1 - \beta_{(0)}^2) - n\theta \cdot \left(\frac{11c_{\mu}}{4} + c_{\mu}^2 + \frac{64c_{\mu}}{4} + \frac{64c_{\mu}}{4}\right) \\
&\geq -\frac{1}{2}n\theta \cdot (1 - \beta_{(0)}^2 + 20c_{\mu}).
\end{aligned} \tag{I.44}$$

The bounds of $\beta^* \boldsymbol{\chi}[\beta]$ can be derive on the event whose expectation is drawn from [Lemma F.2](#) and [\(I.39\)](#) as

$$\mathcal{E}_{\boldsymbol{\chi}} := \left\{ \left\{ \begin{array}{ll} \sigma_i \boldsymbol{\chi}[\beta]_i \geq n\theta \mathcal{S}_{\nu_2 \lambda} [|\beta_i|] - \frac{c_{\mu}n\theta}{p}, & \forall i \in [\pm p] \\ \sigma_i \boldsymbol{\chi}[\beta]_i \leq 6n\theta^2 |\beta_i| + \frac{c_{\mu}n\theta}{p^{3/2}}, & \forall i \neq (0) \end{array} \right\} \right\},$$

then use $\|\beta\|_1 \leq 1 + \frac{\lambda p}{4 \log \theta^{-1}} \leq \frac{\lambda p}{2}$, implies:

$$\begin{aligned}
\beta^* \boldsymbol{\chi}[\beta] &\geq n\theta |\beta_{(0)}| (|\beta_{(0)}| - \nu_2 \lambda) - c_{\mu} \|\beta\|_1 \frac{n\theta}{p} \\
&\geq n\theta \left(\beta_{(0)}^2 - \sqrt{\frac{2}{\pi}} \lambda - \frac{c_{\mu}}{2} \lambda \right) \\
&\geq n\theta \left(\beta_{(0)}^2 - \lambda \right).
\end{aligned} \tag{I.45}$$

Finally via the regional condition $|\beta_{(1)}| \leq \frac{\lambda}{4 \log \theta^{-1}}$, the absolute value of leading correlation

$$\beta_{(0)}^2 \geq \|\beta_\tau\|_2^2 - |\tau| \beta_{(1)}^2 \geq 1 - 2c_\mu - 0.1^2 > 0.9, \quad (\text{I.46})$$

then we collect all above results and obtain:

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_{\ell^1}](\mathbf{a}) \mathbf{v} = \mathbf{v}^* \widetilde{\nabla}^2 \varphi_{\ell^1}(\mathbf{a}) \mathbf{v} - \beta^* \chi[\beta] \geq \left(1.5\beta_{(0)}^2 - 0.5 - \lambda - 20c_\mu\right) n\theta \geq 0.3n\theta, \quad (\text{I.47})$$

with probability at least

$$1 - \underbrace{\mathbb{P}[\mathcal{E}_{\text{cross}}^c]}_{\text{Lemma G.4}} - \underbrace{\mathbb{P}[\mathcal{E}_{\text{pcurv}}^c]}_{\text{Corollary G.3}} - \underbrace{\mathbb{P}[\mathcal{E}_{\|\mathbf{x}\|_2}^c]}_{\text{Lemma D.2}} - \underbrace{\mathbb{P}[\mathcal{E}_\chi^c]}_{\text{Corollary F.4}} \geq 1 - c'/n. \quad (\text{I.48})$$

3. (Identify local minima) Wlog let \mathbf{a}_* be a local minimum where its gradient is zero that is close to \mathbf{a}_0 . The strong convexity (I.47), provides the upper bound on $\|\mathbf{a}_* - \mathbf{a}_0\|_2^2$ via

$$\begin{aligned} \varphi_{\ell^1}(\mathbf{a}_*) &\geq \varphi_{\ell^1}(\mathbf{a}_0) + \langle \mathbf{a}_* - \mathbf{a}_0, \text{grad}[\varphi_{\ell^1}](\mathbf{a}_0) \rangle + \frac{0.3}{2} n\theta \|\mathbf{a}_* - \mathbf{a}_0\|_2^2 \\ \implies \|\text{grad}[\varphi_{\ell^1}](\mathbf{a}_0)\|_2 &\geq 0.15n\theta \|\mathbf{a}_* - \mathbf{a}_0\|_2 \end{aligned} \quad (\text{I.49})$$

Thus we only require to bound the gradient at \mathbf{a}_0 , whose coefficients $\alpha = e_0$ and correlation β has properties $\beta_0 = 1$ and $\|\beta_{\setminus 0}\|_\infty \leq \mu$ hence $\|\beta_{\setminus 0}\|_2 \leq \sqrt{2p}\mu$. Expand the gradient term and condition on \mathcal{E}_χ , since $\mu p^2 \theta^2 \leq \frac{c_\mu}{4}$ and $\theta < \frac{c_\mu}{4\sqrt{p}}$, we can upper bound the gradient at \mathbf{a}_0 as

$$\begin{aligned} \|\text{grad}[\varphi_{\ell^1}](\mathbf{a}_0)\|_2 &= \|\iota^* \mathbf{C}_{\mathbf{a}_0} (\chi[\beta] - \beta^* \chi[\beta] e_0)\|_2 \leq \|\iota^* \mathbf{C}_{\mathbf{a}_0}\|_2 \|\chi[\beta]_{\setminus 0}\|_2 \\ &\leq \sqrt{1 + \mu p} \left(6n\theta^2 \|\beta_{\setminus 0}\|_2 + n\theta \cdot \frac{c_\mu}{p^{3/2}} \cdot \sqrt{2p}\right) \\ &\leq n\theta \sqrt{1 + \mu p} \left(6\mu \sqrt{2p} \cdot \theta + \frac{2c_\mu}{p}\right) \\ &\leq n\theta \left(3c_\mu \mu + 6\mu \cdot \sqrt{2\mu} \cdot p\theta + \frac{2c_\mu}{p} + \frac{2c_\mu \sqrt{\mu}}{\sqrt{p}}\right) \\ &\leq 7\sqrt{c_\mu} n\theta \cdot \max\left\{\mu, \frac{1}{p}\right\}. \end{aligned} \quad (\text{I.50})$$

Thus we conclude that with sufficiently small c_μ :

$$\|\mathbf{a}_* - \mathbf{a}_0\|_2 \leq 50\sqrt{c_\mu} \max\{\mu, p^{-1}\} \leq \frac{1}{2} \max\{\mu, p^{-1}\}. \quad (\text{I.51})$$

and we complete the proof by generalize this result from minima near \mathbf{a}_0 to any of its shifts $s_i[\mathbf{a}_0]$. ■

Similarly, for objective φ_ρ we have

Corollary I.6 (Strong convexity of φ_ρ of near shift). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_ρ with $c_\lambda \in [\frac{1}{5}, \frac{1}{4}]$, then there exists some numerical constant $C, c, c', c'', \bar{c} > 0$ such that if ρ is δ -smoothed ℓ^1 function where $\delta \leq c' \lambda \theta^8 / p^2 \log^2 n$ and $n > Cp^5 \theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ satisfying $|\beta_{(1)}| < \nu_1 \lambda$: for all $\mathbf{v} \in \mathbb{S}^{p-1} \cap \mathbf{a}^\perp$,*

$$\mathbf{v}^* \widetilde{\text{Hess}}[\varphi_\rho](\mathbf{a}) \mathbf{v} > cn\theta; \quad (\text{I.52})$$

furthermore, there exists $\bar{\mathbf{a}}$ as an local minimizer such that

$$\min_\ell \|\bar{\mathbf{a}} - s_\ell[\mathbf{a}_0]\|_2 \leq \frac{1}{2} \max\{\mu, p^{-1}\} \quad (\text{I.53})$$

Proof. The strong convexity (I.52) is derived by combining (I.36) and (H.23) by letting constant multiplier of δ satisfies $c^{1/4} < 10^{-3}c$. On the other hand the local minimizer near solution (I.53) is derived via combining (I.49), (H.21) and utilize both $\theta\sqrt{p} < c_\mu$ and $\mu p^2\theta^2 < c_\mu$ such that:

$$\begin{aligned}\|\text{grad}[\varphi_\rho](\mathbf{a})\|_2 &\leq \|\iota^* \mathbf{C}_{\mathbf{a}_0}\|_2 \left\| \boldsymbol{\chi}[\boldsymbol{\beta}] - \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda^\delta \left[\check{\mathbf{C}}_{\mathbf{y}} \iota \mathbf{a} \right] \right\|_2 + \|\iota^* \mathbf{C}_{\mathbf{a}_0}\|_2 \|\boldsymbol{\chi}[\boldsymbol{\beta}]_{\setminus 0}\|_2 \\ &\leq \sqrt{1 + \mu p} \cdot n\theta^3 + 7\sqrt{c_\mu} n\theta \cdot \max\{\mu, p^{-1}\} \\ &\leq 8n\theta\sqrt{c_\mu} \cdot \max\{\mu, p^{-1}\}\end{aligned}\tag{I.54}$$

I.4 Retraction toward subspace

As in Figure 6, the function value grows in direction away from subspace \mathcal{S}_τ , we will illustrate this phenomenon by proving the negative gradient direction $-\mathbf{g}$ will point toward the subspace \mathcal{S}_τ . To show this, we prove for every coefficients of \mathbf{a} as $\boldsymbol{\alpha}$, there exists coefficients of \mathbf{g} as $\boldsymbol{\zeta}$ satisfies

$$\langle \boldsymbol{\alpha}_{\tau^c}(\mathbf{g}), \boldsymbol{\alpha}_{\tau^c}(\mathbf{a}) \rangle > c \|\boldsymbol{\alpha}_{\tau^c}\|_2 \|\boldsymbol{\zeta}_{\tau^c}\|_2\tag{I.55}$$

whenever $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \in [\frac{\gamma}{2}, \gamma]$. Apparently, the gradient will decrease $d_\alpha(\mathbf{a}, \mathcal{S}_\tau)$, hence being addressed as *retractive toward subspace* \mathcal{S}_τ . This retractive phenomenon is true for gradient of both φ_{ℓ^1} and φ_ρ .

Lemma I.7 (Retraction of φ_{ℓ^1} toward subspace). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda/\sqrt{k}$ in φ_{ℓ^1} with $c_\lambda \in (0, \frac{1}{3}]$, then there exists some numerical constants $C, c, \bar{c} > 0$ such that if $n > Cp^5\theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ such that if*

$$d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \geq \gamma(c_\mu)/2\tag{I.56}$$

then for every $\boldsymbol{\alpha}$ satisfying $\mathbf{a} = \iota^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}$, there exists some $\boldsymbol{\zeta}$ satisfying $\text{grad}[\varphi_{\ell^1}](\mathbf{a}) = \iota^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\zeta}$ that

$$\langle \boldsymbol{\zeta}_{\tau^c}, \boldsymbol{\alpha}_{\tau^c} \rangle \geq \frac{1}{4n\theta} \|\boldsymbol{\zeta}_{\tau^c}\|_2^2.\tag{I.57}$$

Proof. Write $\gamma = \gamma(c_\mu)$ Recall the gradient can be derived as

$$\text{grad}[\varphi_{\ell^1}](\mathbf{a}) = -\mathbf{P}_{\mathbf{a}^\perp} \iota^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\chi}[\boldsymbol{\beta}] = (\mathbf{a}\mathbf{a}^* - \mathbf{I}) \iota^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\chi}[\boldsymbol{\beta}] = \iota^* \mathbf{C}_{\mathbf{a}_0} (\boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}] \boldsymbol{\alpha} - \boldsymbol{\chi}[\boldsymbol{\beta}]),\tag{I.58}$$

for every $\boldsymbol{\alpha}$ satisfies $\mathbf{a} = \iota^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}$. Now via Corollary F.4, condition on the event:

$$\mathcal{E}_\chi := \left\{ \sigma_i \boldsymbol{\chi}[\boldsymbol{\beta}]_i \leq \begin{cases} n\theta \cdot |\boldsymbol{\beta}_i| + \frac{c_\mu n\theta}{p}, & \forall i \in \tau \\ n\theta \cdot |\boldsymbol{\beta}_i| 4\theta |\tau| + \frac{c_\mu n\theta}{p}, & \forall i \in \tau^c \end{cases}, \quad \sigma_i \boldsymbol{\chi}[\boldsymbol{\beta}]_i \geq n\theta \cdot \mathcal{S}_{\sqrt{2/\pi}\lambda} \left[\|\boldsymbol{\beta}_i\| \right] \right\},\tag{I.59}$$

and on this event, utilize Lemma E.5, bounds of $\boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}]$ and $\|\boldsymbol{\chi}[\boldsymbol{\beta}]_{\tau^c}\|_2$ can be derived with $c_\mu < \frac{1}{100}$ as:

$$\boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}] \leq n\theta \left(\|\boldsymbol{\beta}_\tau\|_2^2 + 4\theta |\tau| \|\boldsymbol{\beta}_{\tau^c}\|_2^2 + c_\mu \right) \geq n\theta (1 + c_\mu + 4c_\mu^2 + c_\mu) \leq \frac{3}{2}n\theta\tag{I.60}$$

$$\boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}] \geq n\theta \left(\|\boldsymbol{\beta}_\tau\|_2^2 - \sqrt{2/\pi}\lambda \|\boldsymbol{\beta}_\tau\|_1 - c_\mu \right) \geq n\theta \left(1 - 4c_\mu - \sqrt{2/\pi}c_\lambda - c_\mu \right) \geq \frac{1}{2}n\theta\tag{I.61}$$

$$\|\boldsymbol{\chi}[\boldsymbol{\beta}]_{\tau^c}\|_2 \leq 4n\theta^2 |\tau| \|\boldsymbol{\beta}_{\tau^c}\|_2 + \frac{c_\mu n\theta}{p} \sqrt{p} \leq n\theta (4c_\mu \gamma + c_\mu \gamma) \leq \frac{1}{20}n\theta\gamma.\tag{I.62}$$

Let $\boldsymbol{\alpha}(\mathbf{g}) = \boldsymbol{\beta}^* \boldsymbol{\chi}[\boldsymbol{\beta}] \boldsymbol{\alpha} - \boldsymbol{\chi}[\boldsymbol{\beta}]$, derive

$$\langle \boldsymbol{\alpha}(\mathbf{g})_{\tau^c}, \boldsymbol{\alpha}_{\tau^c} \rangle - \frac{1}{4n\theta} \|\boldsymbol{\alpha}(\mathbf{g})_{\tau^c}\|_2^2$$

$$\begin{aligned}
&= \beta^* \chi[\beta] \|\alpha_{\tau^c}\|_2^2 - \langle \alpha_{\tau^c}, \chi[\beta]_{\tau^c} \rangle \\
&\quad - \frac{1}{4n\theta} \|\beta^* \chi[\beta] \alpha_{\tau^c} - \chi[\beta]_{\tau^c}\|_2^2 \\
&\geq \beta^* \chi[\beta] \|\alpha_{\tau^c}\|_2^2 - \|\alpha_{\tau^c}\|_2 \|\chi[\beta]_{\tau^c}\|_2 \\
&\quad - \frac{1}{2n\theta} |\beta^* \chi[\beta]|^2 \|\alpha_{\tau^c}\|_2^2 - \frac{1}{2n\theta} \|\chi[\beta]_{\tau^c}\|_2^2 \\
&\geq (\beta^* \chi[\beta] - \frac{1}{2n\theta} (\beta^* \chi[\beta])^2) \|\alpha_{\tau^c}\|_2^2 - \frac{1}{20} n\theta\gamma \|\alpha_{\tau^c}\|_2 - \frac{1}{1000} n\theta\gamma^2, \tag{I.63}
\end{aligned}$$

notice that this is a quadratic function of $\beta^* \chi[\beta]$ with negative leading coefficient and zeros at $\{0, 2n\theta\}$, hence (I.63) is minimized when $\beta^* \chi[\beta] = \frac{1}{2}n\theta$. Plugging in,

$$(I.63) \geq \frac{3}{8}n\theta \|\alpha_{\tau^c}\|_2^2 - \frac{1}{20}n\theta\gamma \|\alpha_{\tau^c}\|_2 - \frac{1}{1000}n\theta\gamma^2 \tag{I.64}$$

then again this is a quadratic function of $\|\alpha_{\tau^c}\|_2$ with positive leading coefficient and zeros at $\{0, \frac{8}{60}\gamma\}$, thus (I.64) is minimized at $\|\alpha_{\tau^c}\|_2 = \frac{\gamma}{2}$. Plugging in again,

$$(I.64) \geq \frac{3}{8}n\theta \|\alpha_{\tau^c}\|_2^2 - \frac{1}{20}n\theta\gamma \|\alpha_{\tau^c}\|_2 - \frac{1}{1000}n\theta\gamma^2 \geq (\frac{3}{32} - \frac{1}{80} - \frac{1}{1000})n\theta\gamma^2 > 0 \tag{I.65}$$

which concludes our proof. \blacksquare

As a consequence, we have that

Corollary I.8 (Retraction of φ_ρ toward the subspace). *Suppose that $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ in \mathbb{R}^n , and k, c_μ such that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$. Define $\lambda = c_\lambda / \sqrt{|k|}$ in φ_ρ with $c_\lambda \in (0, \frac{1}{3}]$, then there exists some numerical constants $C, c, c', c'', \bar{c} > 0$ such that if ρ is δ -smoothed ℓ^1 function where $\delta \leq c'' \lambda \theta^8 / p^2 \log^2 n$ and $n > Cp^5 \theta^{-2} \log p$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - c'/n$, for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ such that if*

$$d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \geq \gamma(c_\mu)/2 \tag{I.66}$$

then for every α satisfying $\mathbf{a} = \iota^* \mathbf{C}_{\mathbf{a}_0} \alpha$, there exists some ζ satisfying $\text{grad}[\varphi_\rho](\mathbf{a}) = \iota^* \mathbf{C}_{\mathbf{a}_0} \zeta$ that

$$\langle \zeta_{\tau^c}, \alpha_{\tau^c} \rangle \geq \frac{1}{6n\theta} \|\zeta_{\tau^c}\|_2^2. \tag{I.67}$$

Proof. Write $\gamma = \gamma(c_\mu)$. Define

$$\chi_{\ell^1}[\beta] = \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda[\check{\mathbf{a}} * \mathbf{y}], \quad \chi_\rho[\beta] = \check{\mathbf{C}}_{\mathbf{x}_0} \mathcal{S}_\lambda^\delta[\check{\mathbf{a}} * \mathbf{y}],$$

which, and on event (I.59) and Lemma H.6, we know

$$\beta^* \chi_{\ell^1}[\beta] \leq \frac{3}{2}n\theta, \tag{I.68}$$

$$\|\chi_{\ell^1}[\beta]_{\tau^c}\|_2 \leq \frac{1}{20}n\theta\gamma, \tag{I.69}$$

$$\|\chi_{\ell^1}[\beta] - \chi_\rho[\beta]\|_2 \leq c_1 n\theta^4, \tag{I.70}$$

for some constant $c_1 > 0$. Now given any α satisfies $\mathbf{a} = \iota^* \mathbf{C}_{\mathbf{a}_0} \alpha$, the gradient of both objective can be derived as

$$\begin{aligned}
\text{grad}[\varphi_{\ell^1}](\mathbf{a}) &= -\mathbf{P}_{\mathbf{a}^\perp} \iota^* \mathbf{C}_{\mathbf{a}_0} \text{prox}_{\lambda \|\cdot\|_1}[\check{\mathbf{a}} * \mathbf{y}] = (\mathbf{a}\mathbf{a}^* - \mathbf{I}) \iota^* \mathbf{C}_{\mathbf{a}_0} \chi_{\ell^1}[\beta] \\
&= \iota^* \mathbf{C}_{\mathbf{a}_0} (\beta^* \chi_{\ell^1}[\beta] \alpha - \chi_{\ell^1}[\beta]), \tag{I.71}
\end{aligned}$$

$$\begin{aligned}
\text{grad}[\varphi_\rho](\mathbf{a}) &= -\mathbf{P}_{\mathbf{a}^\perp} \iota^* \mathbf{C}_{\mathbf{a}_0} \text{prox}_{\lambda \rho}[\check{\mathbf{a}} * \mathbf{y}] = (\mathbf{a}\mathbf{a}^* - \mathbf{I}) \iota^* \mathbf{C}_{\mathbf{a}_0} \chi_\rho[\beta] \\
&= \iota^* \mathbf{C}_{\mathbf{a}_0} (\beta^* \chi_\rho[\beta] \alpha - \chi_\rho[\beta]). \tag{I.72}
\end{aligned}$$

In the same spirit, define the coefficient of each gradient vector

$$\zeta_{\ell^1} = \beta^* \chi_{\ell^1}[\beta] \alpha - \chi_{\ell^1}[\beta], \tag{I.73}$$

$$\zeta_\rho = \beta^* \chi_\rho[\beta] \alpha - \chi_\rho[\beta], \quad (\text{I.74})$$

which, by norm inequality from (I.68)-(I.70) and Lemma I.7, we can derive

$$\|\zeta_{\ell^1} - \zeta_\rho\|_2 \leq \|(\mathbf{I} - \alpha\beta^*)(\chi_\rho[\beta] - \chi_{\ell^1}[\beta])\|_2 \leq c_1 n \theta^4, \quad (\text{I.75})$$

$$\|(\zeta_{\ell^1})_{\tau^c}\|_2 \geq |\beta^* \chi_{\ell^1}[\beta]| \|\alpha_{\tau^c}\|_2 - \|\chi_{\ell^1}[\beta]_{\tau^c}\|_2 \geq \frac{1}{5} n \theta \gamma, \quad (\text{I.76})$$

$$\langle (\zeta_{\ell^1})_{\tau^c}, \alpha_{\tau^c} \rangle \geq \frac{1}{4n\theta} \|(\zeta_{\ell^1})_{\tau^c}\|_2^2, \quad (\text{I.77})$$

where the first inequality is derived by observing $(\mathbf{I} - \alpha\beta^*)$ is a projection operator, as such:

$$\begin{aligned} \beta^* \alpha &= \mathbf{a}^* \iota^* \mathbf{C}_{a_0} \alpha = \mathbf{a}^* \mathbf{a} = 1, \\ (\mathbf{I} - \alpha\beta^*)^2 &= \mathbf{I} - 2\alpha\beta^* + \alpha(\beta^* \alpha)\beta^* = \mathbf{I} - \alpha\beta^*. \end{aligned}$$

Now we are ready to derive (I.57):

$$\begin{aligned} \langle (\zeta_\rho)_{\tau^c}, \alpha_{\tau^c} \rangle &\geq \langle (\zeta_{\ell^1})_{\tau^c}, \alpha_{\tau^c} \rangle - \|\alpha_{\tau^c}\|_2 \|\zeta_\rho - \zeta_{\ell^1}\|_2 \\ &\geq \frac{1}{4n\theta} \|(\zeta_{\ell^1})_{\tau^c}\|_2^2 - c_1 n \theta^4 \gamma \\ &\geq \frac{1}{12n\theta} \|(\zeta_{\ell^1})_{\tau^c}\|_2^2 \\ &\quad + \frac{1}{6n\theta} \left(\|(\zeta_\rho)_{\tau^c}\|_2^2 - 2 \|(\zeta_{\ell^1})_{\tau^c}\|_2 \|\zeta_{\ell^1} - \zeta_\rho\|_2 - \|\zeta_{\ell^1} - \zeta_\rho\|_2^2 \right) - c_1 n \theta^4 \gamma \\ &\geq \frac{1}{6n\theta} \|(\zeta_\rho)_{\tau^c}\|_2^2 + \frac{1}{12n\theta} \left(\frac{1}{5} n \theta \gamma \right)^2 - \frac{1}{3n\theta} \left(\frac{1}{5} n \theta \gamma \right) (c_1 n \theta^4) \\ &\quad - \frac{1}{6n\theta} (c_1 n \theta^4)^2 - c_1 n \theta^4 \gamma \\ &\geq \frac{1}{6n\theta} \|(\zeta_\rho)_{\tau^c}\|_2^2. \end{aligned} \quad (\text{I.78})$$

where the last inequality is true since $\theta^3 \ll \gamma$. ■

I.5 Proof of Theorem B.1

By collecting result from above, we are ready to prove the acclaimed geometric result in Theorem B.1. It guarantees that for every \mathbf{a} near \mathcal{S}_τ , either one of the following is true

$$\lambda_{\min}(\text{Hess}[\varphi_\rho](\mathbf{a})) \leq -c_1 n \theta \lambda, \quad (\text{I.79})$$

$$\langle \sigma_{(0)} \iota^* s_{(0)}[\mathbf{a}_0], -\text{grad}[\varphi_\rho](\mathbf{a}) \rangle \geq c_2 n \theta (\log^{-2} \theta^{-1}) \lambda^2, \quad (\text{I.80})$$

$$\text{Hess}[\varphi_\rho](\mathbf{a}) \succ c_3 n \theta \cdot \mathbf{P}_{\mathbf{a}^\perp}, \quad (\text{I.81})$$

all local minimizer $\bar{\mathbf{a}}$ satisfies for some $\mathbf{a}_* \in \{\pm \iota^* s_\ell[\mathbf{a}] \mid \ell \in [\pm p_0]\}$,

$$\|\bar{\mathbf{a}} - \mathbf{a}_*\|_2 \leq c_4 \sqrt{c_\mu} \max\{\mu, p_0^{-1}\}, \quad (\text{I.82})$$

and whenever $\frac{\gamma}{2} \leq d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \leq \gamma$, coefficient of \mathbf{a} and its gradient \mathbf{g} , α , written as ζ , satisfies

$$\langle \zeta_{\tau^c}, \alpha_{\tau^c} \rangle \geq \frac{c_5}{n\theta} \|\zeta_{\tau^c}\|_2^2. \quad (\text{I.83})$$

To connect the geometric results introduced in Lemma I.1, Lemma I.3, Lemma I.5 and Lemma I.7, we are only required to prove the required signal condition claimed in Theorem B.1 is necessary from Definition E.1. In particular, when the subspace dimension $|\tau| \leq 4p_0\theta$. On top of that, we are also required to show the chosen smooth parameter δ in the pseudo-Huber penalty $\rho(x) = \sqrt{x^2 + \delta^2}$ approximate $|x|$ sufficiently well, hence results of Corollary I.2, Corollary I.4, Corollary I.6 and Corollary I.8 also holds.

Proof. Firstly we will show when largest solution subspace dimension $k = 4p_0\theta$, the signal condition of [Definition E.1](#) will be satisfied. Recall that the signal condition of [Theorem B.1](#) requests

$$\frac{2}{p_0 \log^2 p_0} \leq \theta \leq \frac{c}{(p_0 \sqrt{\mu} + \sqrt{p_0}) \log^2 p_0}, \quad (\text{I.84})$$

since $p = 3p_0 - 2$, this implies the lower bounds for sparsity θ as:

$$\theta \geq \frac{1}{2p_0 \left(\frac{1}{2} \log p_0\right)^2} \geq \frac{1}{p \log^2 \theta^{-1}}; \quad (\text{I.85})$$

the upper bound of θ via $\theta \sqrt{p_0} \log^2 p_0 \leq c$:

$$\theta \leq \frac{9c}{\sqrt{p_0} (3 \log p_0)^2} \leq \frac{16c}{\sqrt{p} \log^2 \theta^{-1}}, \quad \theta \leq \frac{4c^2}{k \log^4 p_0} \leq \frac{36c^2}{k (3 \log p_0)^2} \leq \frac{36c^2}{k \log^2 \theta^{-1}}; \quad (\text{I.86})$$

and the upper bound for coherence μ as:

$$\begin{aligned} \mu \max \{k^2, (p\theta)^2\} \log^2 \theta^{-1} &\leq \mu \max \{16(p_0\theta)^2, 9(p_0\theta)^2\} \log^2 \theta^{-1} \\ &\leq 16 (\sqrt{\mu} p_0 \theta)^2 \log^2 p_0 \leq 16c. \end{aligned} \quad (\text{I.87})$$

Therefore [Definition E.1](#) holds if $\max \{16c, 36c^2\} \leq c_\mu/4$ via (I.85)-(I.87).

Furthermore, we know from lemma assumption all interested \mathbf{a} are near subspace \mathcal{S}_τ by

$$\begin{aligned} d_\alpha(\mathbf{a}, \mathcal{S}_\tau) &\leq \frac{c}{\sqrt{p_0} \log^2 \theta^{-1}} \cdot \min \left\{ \frac{1}{\sqrt{\theta}}, \frac{1}{\sqrt{\mu}}, \frac{1}{\mu (p_0\theta)^{3/2}} \right\} \\ &\leq \frac{c}{\log^2 \theta^{-1}} \min \left\{ \frac{2}{\sqrt{k}}, \frac{1}{\sqrt{p_0\mu}}, \frac{4}{\mu p_0 \sqrt{\theta} k} \right\} \leq \gamma \end{aligned} \quad (\text{I.88})$$

where γ is defined in [Definition E.3](#) of widened subspace $\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$.

Lastly, the pseudo-Huber function $\rho(x) = \sqrt{x^2 + \delta^2}$ is an ℓ^1 smoothed sparse surrogate defined in [Definition H.2](#), by observing that it is convex, smooth, even, whose second order derivative (according to [Table 1](#)) $\nabla^2 \rho(x) = \frac{\delta^2}{(x^2 + \delta^2)^{3/2}}$ is monotone decreasing in $|x|$. More importantly

$$\sup_{x \in \mathbb{R}} |\rho(x) - |x|| = |\rho(0) - |0|| = \delta. \quad (\text{I.89})$$

Hence, by choosing $\delta \leq \frac{c'^4 \theta^8}{p^2 \log^2 n} \lambda$, for some sufficiently small constant c' and letting $\lambda = 0.2\sqrt{k} = 0.1/\sqrt{p_0\theta}$ in φ_ρ . We obtain the geometrical results in [Corollary I.2](#) when $|\beta_{(1)}| \geq \frac{4}{5} |\beta_{(0)}|$, [Corollary I.4](#) when $\frac{4}{5} |\beta_{(0)}| \geq |\beta_{(1)}| \geq \frac{\lambda}{4 \log^2 \theta^{-1}}$ and [Corollary I.6](#) when $\frac{\lambda}{4 \log^2 \theta^{-1}} \geq |\beta_{(1)}|$, and the retraction result in [Corollary I.8](#). ■

J Analysis of algorithm — minimization within widened subspace

In this section, we prove convergence of the first part of our algorithm—minimization of φ_ρ near \mathcal{S}_τ . We begin by proving the initialization method guarantees that $\mathbf{a}^{(0)}$ is near \mathcal{S}_τ , in the sense that

$$d_\alpha(\mathbf{a}^{(0)}, \mathcal{S}_\tau) \leq \gamma, \quad (\text{J.1})$$

where the distance d_α is defined in (B.15). We then demonstrate that small-stepping curvilinear search converges to a desired local minimum of φ_ρ at rate $O(1/k)$, where k is the iteration number. To do this, it is important to utilize(i) the *retractive* property to show that the iterates stay near \mathcal{S}_τ and (ii) the geometric properties of φ_ρ near \mathcal{S}_τ .

J.1 Initialization near subspace

The following lemma shows that the initialization $\mathbf{a}^{(0)} = \mathbf{P}_{\mathbb{S}^{p-1}} [\nabla \varphi_{\ell^1}(\mathbf{a}^{(-1)})]$, where

$$\mathbf{a}^{(-1)} = \mathbf{P}_{\mathbb{S}^{p-1}} \left[\sum_{\ell \in \tau} \mathbf{x}_0 \ell \boldsymbol{\iota}_{p_0}^* s_\ell[\mathbf{a}_0] \right], \quad (\text{J.2})$$

and is very close to the subspace \mathcal{S}_τ :

Lemma J.1 (Initialization from a piece of data). *Let $\bar{\mathbf{x}} \in \mathbb{R}^{2p_0-1}$ indexed by $[\pm p_0]$, with $\bar{\mathbf{x}}_i \sim_{\text{i.i.d.}} \text{BG}(\theta)$. Define $\bar{\mathbf{y}} = \bar{\mathbf{x}} * \mathbf{a}_0$, and $\mathbf{a}^{(0)}$ as*

$$\mathbf{a}^{(0)} = -\mathbf{P}_{\mathbb{S}^{p-1}} \nabla \varphi_{\ell^1} \left(\mathbf{P}_{\mathbb{S}^{p-1}} \left[\mathbf{0}^{p_0-1}; [\bar{\mathbf{y}}_0; \dots; \bar{\mathbf{y}}_{p_0-1}]; \mathbf{0}^{p_0-1} \right] \right), \quad (\text{J.3})$$

with $\lambda = 0.2/\sqrt{p\theta}$ in φ_1 . Set $\tau = \text{supp}(\bar{\mathbf{x}})$. Suppose that $(\mathbf{a}_0, \theta, k)$ satisfies the sparsity-coherence condition $\text{SCC}(c_\mu)$ and \mathbf{a}_0 satisfies $\max_{i \neq j} |\langle \boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0], \boldsymbol{\iota}_{p_0}^* s_j[\mathbf{a}_0] \rangle| \leq \mu$. Then there exists some constant $c, \bar{c} > 0$ such that if $p_0\theta > 1000c$ and $c_\mu \leq \bar{c}$, then with probability at least $1 - 1/c$, we have

$$d_\alpha \left(\mathbf{a}^{(0)}, \mathcal{S}_\tau \right) \leq \frac{c_\mu}{4 \log^2 \theta^{-1}} \min \left\{ \frac{1}{\sqrt{|\tau|}}, \frac{1}{\sqrt{\mu p}}, \frac{1}{\mu p \sqrt{\theta} |\tau|} \right\}. \quad (\text{J.4})$$

Proof. 1. (Distance to \mathcal{S}_τ from $\mathbf{a}^{(0)}$) Let $\eta = \|\boldsymbol{\iota}_{p_0}^*(\mathbf{a}_0 * \mathbf{x})\|_2 = \|\boldsymbol{\iota}_{p_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}\|_2$ and $\gamma = \gamma(c_\mu)$, as in (J.4). Expand the expression of $\mathbf{a}^{(0)}$ from (J.3) we have

$$\begin{aligned} \mathbf{a}^{(0)} &= \mathbf{P}_{\mathbb{S}^{p-1}} \boldsymbol{\iota}^* \check{\mathbf{C}}_{\mathbf{y}} \mathcal{S}_\lambda \left[\check{\mathbf{C}}_{\mathbf{y}} \boldsymbol{\iota}_{p_0} \mathbf{P}_{\mathbb{S}^{p_0-1}} \boldsymbol{\iota}_{p_0}^*(\mathbf{a}_0 * \mathbf{x}) \right] \\ &= \mathbf{P}_{\mathbb{S}^{p-1}} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\chi} \left[\frac{1}{\eta} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}_{p_0} \boldsymbol{\iota}_{p_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x} \right] \end{aligned} \quad (\text{J.5})$$

To relate $\mathbf{a}^{(0)}$ to its coefficient, introduce the truncated autocorrelation matrix $\widetilde{\mathbf{M}} = \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota}_{p_0} \boldsymbol{\iota}_{p_0}^* \mathbf{C}_{\mathbf{a}_0}$, define $\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}$ as

$$\tilde{\boldsymbol{\beta}} = \frac{1}{\eta} \widetilde{\mathbf{M}} \mathbf{x}, \quad \tilde{\boldsymbol{\alpha}} = \boldsymbol{\chi} \left[\frac{1}{\eta} \widetilde{\mathbf{M}} \mathbf{x} \right] = \boldsymbol{\chi}[\tilde{\boldsymbol{\beta}}] \quad (\text{J.6})$$

and note that $\widetilde{\mathbf{M}}$ is bounded entrywise as

$$\left| \widetilde{\mathbf{M}}_{ij} \right| \leq \begin{cases} 1 & i = j \in [-p_0 + 1, p_0 - 1] \\ \mu & i \neq j \in [-p_0 + 1, p_0 - 1], |i - j| < p_0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{J.7})$$

From (J.5), we can write $\mathbf{a}^{(0)} = \mathbf{P}_{\mathbb{S}^{p-1}} \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \tilde{\boldsymbol{\alpha}}$, meaning that the normalized version of $\tilde{\boldsymbol{\alpha}}$ is a valid coefficient vector for $\mathbf{a}^{(0)}$. Let $\tau^c = [\pm 2p_0] \setminus \tau$. The distance d_α to subspace \mathcal{S}_τ (B.15) is upper bounded as

$$\begin{aligned} d_\alpha(\mathbf{a}^{(0)}, \mathcal{S}_\tau) &\leq \frac{\|\tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2}{\|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \tilde{\boldsymbol{\alpha}}\|_2} \leq \frac{\|\tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2}{\|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \tilde{\boldsymbol{\alpha}}_\tau\|_2 - \|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2} \\ &\leq \frac{\|\tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2}{\sqrt{1 - \mu |\tau|} \|\tilde{\boldsymbol{\alpha}}_\tau\|_2 - \sqrt{1 + \mu p} \|\tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2} \end{aligned}$$

where the last inequality is derived with Lemma E.4. Therefore, it is sufficient to show

$$\left(1 + \gamma \sqrt{1 + \mu p} \right) \|\tilde{\boldsymbol{\alpha}}_{\tau^c}\|_2 \leq \gamma \sqrt{1 - \mu |\tau|} \|\tilde{\boldsymbol{\alpha}}_\tau\|_2 \quad (\text{J.8})$$

to complete the proof that $d_\alpha(\mathbf{a}^{(0)}, \mathcal{S}_\tau) \leq \gamma$.

2. (Bound η) Condition on the following two events

$$\mathcal{E}_\tau := \{|\tau| < 4p_0\theta\}, \quad \mathcal{E}_{\|\mathbf{x}\|_2} := \left\{ \sqrt{p_0\theta} \leq \|\mathbf{x}\|_2 \leq \sqrt{3p_0\theta} \right\} \quad (\text{J.9})$$

and utilize μ bound from [Lemma E.5](#) such that $\mu|\tau| < 0.1$. An upper bound on η can be obtained using properties of $\widetilde{\mathbf{M}}$ of [\(J.7\)](#):

$$\eta = \|\boldsymbol{\iota}_{p_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}\|_2 \leq \|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}\|_2 \leq \sqrt{1 + \mu|\tau|} \|\mathbf{x}\|_2 \leq 2\sqrt{p_0\theta} \quad (\text{J.10})$$

To lower bound η , use $\eta^2 = \mathbf{g}^* \mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau \mathbf{g}$ where \mathbf{g} is the standard Gaussian vector. Observe the submatrix of $\widetilde{\mathbf{M}}$ is diagonal dominant:

$$\begin{cases} \widetilde{\mathbf{M}}_{ii} = \|\boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0]\|_2^2 \in [0, 1] \\ \text{tr}(\widetilde{\mathbf{M}}) = \sum_{i \in [\pm p_0]} \|\boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0]\|_2^2 = \|\mathbf{a}_0\|_2^2 + \sum_{i=1}^{p_0-1} \left(\|\boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0]\|_2^2 + \|\boldsymbol{\iota}_{p_0}^* s_{i-p_0}[\mathbf{a}_0]\|_2^2 \right) = p_0 \end{cases} \quad (\text{J.11})$$

Write $\mathbf{x} = \mathbf{g} \circ \mathbf{w}$ where \mathbf{w} and \mathbf{g} are Bernoulli and Gaussian vector respectively with $\text{supp}(\mathbf{w}) = \tau$, then the trace of $\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau$ can be written as sum of independent r.v.s as:

$$\text{tr}(\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau) = \sum_{i \in [\pm p_0]} w_i \|\boldsymbol{\iota}_{p_0}^* s_i[\mathbf{a}_0]\|_2^2,$$

Bernstein inequality [Lemma N.4](#) and [\(J.11\)](#) gives

$$\begin{aligned} \mathbb{P} \left[\text{tr}(\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau) < \frac{3p_0\theta}{4} \right] &\leq \mathbb{P} \left[\text{tr}(\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau) - p_0\theta \leq -\frac{p_0\theta}{4} \right] \\ &\leq 2 \exp \left(\frac{-(p_0\theta/4)^2}{2p_0\theta + p_0\theta/2} \right) \leq 2 \exp \left(\frac{-p_0\theta}{40} \right), \end{aligned} \quad (\text{J.12})$$

thus condition on ω satisfies $\text{tr}(\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau) \geq 3p_0\theta/4$ and \mathcal{E}_τ , expectation η^2 has lower bound

$$\mathbb{E}_{\mathbf{g}|\mathbf{w}} \eta^2 = \mathbb{E}_{\mathbf{g}|\mathbf{w}} [\mathbf{g}^* \mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau \mathbf{g}] = \text{tr}(\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau) \geq \frac{3p_0\theta}{4}$$

then apply Bernstein inequality again by first writing svd of $\mathbf{P}_\tau \widetilde{\mathbf{M}} \mathbf{P}_\tau = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^*$ with $\boldsymbol{\Sigma}$ being rank $|\tau| < 4p_0\theta$ and square orthobasis \mathbf{U} . Let $\mathbf{g}' = \mathbf{U}^* \mathbf{g}$, then \mathbf{g}' is standard i.i.d. Gaussian vector, provides alternative expression $\eta^2 < \sum_{i=1}^{4p_0\theta} g_i'^2 \sigma_i$ where $\sigma_i \leq 1 + \mu|\tau| \leq 1.1$. We obtain probability of η^2 to be small as

$$\begin{aligned} \mathbb{P}_{\mathbf{g}|\mathbf{w}} \left[\eta^2 < \frac{p_0\theta}{2} \right] &\leq \mathbb{P}_{\mathbf{g}|\mathbf{w}} \left[\eta^2 - \mathbb{E}_{\mathbf{g}|\mathbf{w}} \eta^2 < -\frac{p_0\theta}{4} \right] \\ &\leq 2 \exp \left(\frac{-(p_0\theta/4)^2}{2(1 + \mu|\tau|)(12p_0\theta + p_0\theta/2)} \right) \leq 2 \exp \left(\frac{-p_0\theta}{440} \right) \end{aligned} \quad (\text{J.13})$$

by applying moment bounds $(\sigma^2, R) = (12p_0\theta(1 + \mu|\tau|), 2(1 + \mu|\tau|))$. We thereby define event

$$\mathcal{E}_\eta = \left\{ \sqrt{p_0\theta/2} \leq \eta \leq 2\sqrt{p_0\theta} \right\}, \quad (\text{J.14})$$

which holds w.h.p. based on [\(J.9\)](#), [\(J.12\)](#) and [\(J.13\)](#).

3. (**Bound $\tilde{\alpha}$**) Condition on $\mathcal{E}_\eta \cap \mathcal{E}_{\|\mathbf{x}\|_2} \cap \mathcal{E}_\tau$. Use definition $\tilde{\beta} = \frac{1}{\eta} \tilde{M} \mathbf{x}$ from (J.6), and properties of \tilde{M} from (J.7) we can obtain:

$$\begin{cases} \|\tilde{\beta}_{\tau^c}\|_2 \leq \frac{1}{\eta} \|\mathbf{L}_{\tau^c}^* \tilde{M} \mathbf{L}_{\tau}\|_2 \|\mathbf{x}\|_2 \leq \frac{\mu \sqrt{p_0 |\tau|}}{\sqrt{p_0 \theta / 2}} \cdot \sqrt{3 p_0 \theta} \leq 3 \mu \sqrt{p_0 |\tau|} \\ \|\tilde{\beta}_\tau\|_2 \geq \frac{1}{\eta} \|\mathbf{L}_\tau^* \tilde{M} \mathbf{L}_\tau\|_2 \|\mathbf{x}\|_2 \geq \frac{\sqrt{1-\mu} |\tau|}{2 \sqrt{p_0 \theta}} \cdot \sqrt{p_0 \theta} \geq 0.45 \end{cases} \quad (\text{J.15})$$

Use definition $\|\tilde{\alpha}\|_2 = \|\chi[\tilde{\beta}]\|_2$, condition on event

$$\mathcal{E}_\chi := \left\{ \begin{cases} \sigma_i \chi[\beta]_i \geq n \theta \mathcal{S}_{\nu_2 \lambda} [|\beta_i|] - \frac{c_\mu n \theta}{p}, & \forall i \in \tau \\ \sigma_i \chi[\beta]_i \leq 4 n \theta^2 |\tau| |\beta_i| + \frac{c_\mu n \theta}{p}, & \forall i \in \tau^c \end{cases} \right\},$$

also from Definition E.1 we have $\mu (p\theta)^{1/2} |\tau|^{3/2} < \frac{c_\mu}{4 \log^2 \theta^{-1}}$ and from lemma assumption $\lambda = \frac{1}{5 \sqrt{p\theta}}$, provides bounds of $\|\tilde{\alpha}\|_2$ via triangle inequality as:

$$\begin{cases} \|\tilde{\alpha}_{\tau^c}\|_2 \leq 4 n \theta^2 |\tau| \cdot \|\tilde{\beta}_{\tau^c}\|_2 + \frac{c_\mu n \theta}{p} \cdot \sqrt{2 p_0} \leq 3 c_\mu n \theta \left(\frac{\sqrt{\theta}}{\log^2 \theta^{-1}} + \frac{c_\mu}{p} \right) \\ \|\tilde{\alpha}_\tau\|_2 \geq n \theta \left(\|\tilde{\beta}_\tau\|_2 - \nu_2 \lambda \sqrt{|\tau|} - \frac{c_\mu}{p} \sqrt{|\tau|} \right) \geq n \theta \left(0.45 - \sqrt{\frac{2}{\pi}} \cdot \frac{1}{5} - c_\mu \right) \geq 0.2 n \theta \end{cases}, \quad (\text{J.16})$$

since both $\theta |\tau|, \mu p \theta |\tau| < c_\mu$, we have

$$\begin{cases} \sqrt{1 + \mu p} \|\tilde{\alpha}_{\tau^c}\|_2 \leq 3 c_\mu n \theta \sqrt{1 + \mu p} \left(\sqrt{\theta} + p^{-1} \right) \leq 6 c_\mu n \theta \\ \|\tilde{\alpha}_{\tau^c}\|_2 \leq \frac{6 c_\mu^{3/2} n \theta}{\log^2 \theta^{-1}} \min \left\{ \frac{1}{\sqrt{|\tau|}}, \frac{1}{\sqrt{\mu p}}, \frac{1}{\mu p \sqrt{\theta} |\tau|} \right\} \leq 24 \sqrt{c_\mu} n \theta \gamma \end{cases},$$

which satisfies (J.8), since $\mu |\tau| < c_\mu < \frac{1}{1000}$,

$$\begin{aligned} (1 + \gamma \sqrt{1 + \mu p}) \|\tilde{\alpha}_{\tau^c}\|_2 &\leq (24 \sqrt{c_\mu} + 6 c_\mu) n \theta \gamma \leq 0.1 n \theta \gamma \\ &\leq \gamma \sqrt{1 - \mu |\tau|} \|\tilde{\alpha}_\tau\|_2. \end{aligned} \quad (\text{J.17})$$

Finally, given $p_0 \theta > 1000 c$, this result holds with probability at least

$$\begin{aligned} 1 - \underbrace{\mathbb{P}[\mathcal{E}_\tau^c]}_{\text{Lemma D.1}} - \underbrace{\mathbb{P}[\mathcal{E}_{\|\mathbf{x}\|_2}^c]}_{\text{Lemma D.2}} - \underbrace{\mathbb{P}[\mathcal{E}_\eta^c]}_{\text{(J.14)}} - \underbrace{\mathbb{P}[\mathcal{E}_\chi^c]}_{\text{Corollary F.4}} \\ \geq 1 - \frac{2}{p_0 \theta} - \frac{1}{n} - 4 \exp\left(\frac{-p_0 \theta}{440}\right) \geq 1 - \frac{1}{c} \end{aligned} \quad (\text{J.18})$$

■

J.2 Minimization near subspace (Proof of Theorem C.1)

Before we start the proof of theorem, writing $\mathbf{g} = \text{grad}[\varphi_\rho](\mathbf{a})$ and $\mathbf{H} = \text{Hess}[\varphi_\rho](\mathbf{a})$, we will first restate the results of Theorem B.1 in simplified terms. The theorem shows that for any $\mathbf{a} \in \mathbb{S}^{p-1}$ whose distance to subspace $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \leq \gamma$, then at least one of the the following statement hold:

$$\|\mathbf{g}\|_2 \geq \eta_g \quad (\text{J.19})$$

$$\lambda_{\min}(\mathbf{H}) \leq -\eta_v \quad (\text{J.20})$$

$$\mathbf{H} \succ \eta_c \cdot \mathbf{P}_{\mathbf{a}^\perp}. \quad (\text{J.21})$$

Furthermore, φ_ρ is retractive near \mathcal{S}_τ : wherever $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \geq \frac{\gamma}{2}$, writing $\alpha(\mathbf{a}), \alpha(\mathbf{g})$ to be the coefficient of \mathbf{a}, \mathbf{g} , we have

$$\langle \alpha(\mathbf{a})_{\tau^c}, \alpha(\mathbf{g})_{\tau^c} \rangle \geq \eta_r \|\alpha(\mathbf{g})_{\tau^c}\|_2. \quad (\text{J.22})$$

Also, the the gradient, Hessian and the third order derivative are all bounded as follows:

Remark J.2. *With high probability, for every \mathbf{a} whose $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) < \gamma$, its $\max\{\|\mathbf{g}\|_2, \|\mathbf{H}\|_2, \|\nabla \mathbf{H}\|_2\} \leq \bar{\eta} = \text{poly}(n, p)$.*

We state [Remark J.2](#) without explicit proof since its derivation is similar to the proof in [Theorem B.1](#).

We prove that if the negative curvature direction $-\mathbf{v}$ is chosen to be the least eigenvector with $\mathbf{v}^* \mathbf{H} \mathbf{v} < -\eta_v$ and $\mathbf{v}^* \mathbf{g}$ (if cannot, let $\mathbf{v} = \mathbf{0}$), then the iterates

$$\mathbf{a}^{(k+1)} = \mathbf{P}_{\mathbb{S}^{p-1}} \left[\mathbf{a}^{(k)} - t\mathbf{g}^{(k)} - t^2\mathbf{v}^{(k)} \right] \quad (\text{J.23})$$

converges toward the minimizer $\bar{\mathbf{a}}$ in ℓ^2 -norm with rate $O(1/k)$. Notice that here all $\eta_g, \eta_v, \eta_c, \eta_r, \bar{\eta}$ are all greater than 0 and are rational functions of the dimension parameters n, p .

Finally, we should note that \mathbf{a}_0 being μ -truncated shift coherent implies that \mathbf{a}_0 is at at most 2μ -shift coherent. Hence we utilize the usual incoherence condition in the proof.

Proof. Notice that when \mathbf{a} is in the region near some signed shift $\bar{\mathbf{a}}$ of \mathbf{a}_0 , the function φ_ρ is strongly convex, and the iterates coincide with the Riemannian gradient method, which converges at a linear rate. Indeed, if for all k larger than some \bar{k} , $\mathbf{a}^{(k)}$ is in this region, then $\|\mathbf{a}^{(k)} - \bar{\mathbf{a}}\|_2 \leq (1 - t\eta_c)^{-(k-\bar{k})} \|\mathbf{a}^{(\bar{k})} - \bar{\mathbf{a}}\|_2$ [[AMS09](#)](Theorem 4.5.6) where the step size $t = \Omega(1/n\theta)$ hence $t\eta_c = \Omega(1)$. We will argue that the iterates $\mathbf{a}^{(k)}$ remain close to the subspace \mathcal{S}_τ and that after $\bar{k} = \text{poly}(n, p)$ iterations they indeed remain in the strongly convex region around some $\bar{\mathbf{a}}$.

1. (Existence of Armijo steplength). First, we show there exists a nontrivial step size t at every iteration, in the sense that for all $\mathbf{a} \in \mathbb{S}^{p-1}$, there exists $T > 0$ such that for all $t \in (0, T)$, the Armijo step condition ([C.11](#)) is satisfied. Note that since φ_ρ is a smooth function, $\mathbf{a} \rightarrow \varphi_\rho \circ \mathbf{P}_{\mathbb{S}^{p-1}}(\mathbf{a})$ admits a version of Taylor's theorem (see also [[AMS09](#)](Section 7.1.3)): for any $\boldsymbol{\xi} \perp \mathbf{a}$, writing $\mathbf{a}^+ = \mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{a} + \boldsymbol{\xi}]$,

$$|\varphi_\rho(\mathbf{a}^+) - (\varphi_\rho(\mathbf{a}) + \langle \text{grad}[\varphi_\rho](\mathbf{a}), \boldsymbol{\xi} \rangle + \frac{1}{2} \boldsymbol{\xi}^* \text{Hess}[\varphi_\rho](\mathbf{a}) \boldsymbol{\xi})| \leq \bar{\eta} \|\boldsymbol{\xi}\|_2^3, \quad (\text{J.24})$$

using $\|\nabla \mathbf{H}\|_2 \leq \bar{\eta}$. Now, let $\boldsymbol{\xi} = -t\mathbf{g} - t^2\mathbf{v}$ as in the iterates ([C.10](#)). Suppose the Armijo step condition ([C.11](#)) does not hold, so

$$\varphi_\rho(\mathbf{a}^+) > \varphi_\rho(\mathbf{a}) - \frac{1}{2} \left(t \|\mathbf{g}\|_2^2 + \frac{1}{2} t^4 \eta_v \|\mathbf{v}\|_2^2 \right). \quad (\text{J.25})$$

Since $\mathbf{g}^* \mathbf{v} \geq 0$ and $\mathbf{v}^* \mathbf{H} \mathbf{v} \leq -\eta_v \|\mathbf{v}\|_2^2$ or $\mathbf{v} = \mathbf{0}$, using $\|\mathbf{a} + \mathbf{b}\|_2^3 \leq 4 \|\mathbf{a}\|_2^3 + 4 \|\mathbf{b}\|_2^3$ (Hölder's inequality) and $\|\mathbf{H}\|_2 < \bar{\eta}$, we can derive

$$\begin{aligned} & \langle \mathbf{g}, -t\mathbf{g} - t^2\mathbf{v} \rangle + \frac{1}{2} (t\mathbf{g} + t^2\mathbf{v})^* \mathbf{H} (t\mathbf{g} + t^2\mathbf{v}) \\ & \quad + c \|\mathbf{g} + t^2\mathbf{v}\|_2^3 > -\frac{1}{2} \left(t \|\mathbf{g}\|_2^2 + \frac{1}{2} t^4 \eta_v \|\mathbf{v}\|_2^2 \right) \\ \implies & -\frac{1}{2} t \|\mathbf{g}\|_2^2 + \frac{1}{2} t^2 \mathbf{g}^* \mathbf{H} \mathbf{g} + t^3 \mathbf{v}^* \mathbf{H} \mathbf{g} \\ & \quad - \frac{1}{4} t^4 \eta_v \|\mathbf{v}\|_2^2 + 4\bar{\eta} t^3 \|\mathbf{g}\|_2^3 + 4\bar{\eta} t^6 \|\mathbf{v}\|_2^3 > 0 \\ \implies & -\frac{1}{2} t \|\mathbf{g}\|_2^2 + t^2 \left(\frac{1}{2} \bar{\eta} \|\mathbf{g}\|_2^2 + t\bar{\eta} \|\mathbf{v}\|_2 \|\mathbf{g}\|_2 + 4\bar{\eta} t \|\mathbf{g}\|_2^3 \right) \\ & \quad - \frac{1}{4} t^4 \eta_v \|\mathbf{v}\|_2^2 + 4\bar{\eta} t^6 \|\mathbf{v}\|_2^3 > 0. \end{aligned} \quad (\text{J.26})$$

If

$$t < T = \min \left\{ \frac{\|\mathbf{g}\|_2}{\bar{\eta} \|\mathbf{g}\|_2 + 2\bar{\eta}t \|\mathbf{v}\|_2 + 8\bar{\eta}t \|\mathbf{g}\|_2^2}, \sqrt{\frac{\eta_v}{16\bar{\eta} \|\mathbf{v}\|_2}} \right\}, \quad (\text{J.27})$$

then (J.26) < 0 contradicting (J.25). Using our bounds on $\|\mathbf{g}\|_2$, $\bar{\eta}$, η_v and $\|\mathbf{v}\|$, it follows that T is lower bounded by a polynomial $\text{poly}(n^{-1}, p^{-1})$.

2.(Bounds on $d_\alpha(\mathbf{g}, \mathcal{S}_\tau)$, $d_\alpha(\mathbf{v}, \mathcal{S}_\tau)$) We will show there are numerical constants c_g, c_v such that

$$d_\alpha(\mathbf{g}, \mathcal{S}_\tau) \leq c_g n \theta \gamma \quad \text{and} \quad d_\alpha(\mathbf{v}, \mathcal{S}_\tau) \leq c_v n \theta p. \quad (\text{J.28})$$

Define

$$\chi_{\ell^1}[\boldsymbol{\beta}] = \check{\mathbf{C}}_{x_0} \text{prox}_{\lambda \ell^1}[\check{\mathbf{a}} * \mathbf{y}], \quad \chi_\rho[\boldsymbol{\beta}] = \check{\mathbf{C}}_{x_0} \text{prox}_{\lambda \rho}[\check{\mathbf{a}} * \mathbf{y}],$$

then the gradient can be written as (I.58)

$$\text{grad}[\varphi_{\ell^1}](\mathbf{a}) = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} (\boldsymbol{\beta}^* \chi_{\ell^1}[\boldsymbol{\beta}] \boldsymbol{\alpha} - \chi_{\ell^1}[\boldsymbol{\beta}]), \quad (\text{J.29})$$

$$\text{grad}[\varphi_\rho](\mathbf{a}) = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} (\boldsymbol{\beta}^* \chi_\rho[\boldsymbol{\beta}] \boldsymbol{\alpha} - \chi_\rho[\boldsymbol{\beta}]). \quad (\text{J.30})$$

Use the following inequalities:

$$\begin{aligned} \frac{1}{2} n \theta &\leq |\boldsymbol{\beta}^* \chi_{\ell^1}[\boldsymbol{\beta}]| \leq \frac{3}{2} n \theta, \\ \|\chi_{\ell^1}[\boldsymbol{\beta}]_{\tau^c}\|_2 &\leq \frac{1}{20} n \theta \gamma, \\ \|\mathbf{I} - \boldsymbol{\alpha} \boldsymbol{\beta}^*\|_2 &\leq 4\sqrt{p}, \\ \|\chi_{\ell^1}[\boldsymbol{\beta}] - \chi_\rho[\boldsymbol{\beta}]\|_2 &\leq n \theta^4, \end{aligned}$$

where the first and second bounds of $\chi_{\ell^1}[\boldsymbol{\beta}]$ based on event (I.59); the third by observing $\|\boldsymbol{\alpha}\|_2 \leq 2$ and $\|\boldsymbol{\beta}\|_2 \leq 2 + c_\mu \sqrt{p}$; the last from (H.21) of Lemma H.6 when δ is sufficiently small. Hence, by definition of $d_\alpha(\cdot, \mathcal{S}_\tau)$ (B.15) and knowing \mathbf{a} is close to subspace $\|\boldsymbol{\alpha}_{\tau^c}\|_2 \leq \gamma$, via triangle inequality, we get

$$\begin{aligned} d_\alpha(\mathbf{g}, \mathcal{S}_\tau) &\leq d_\alpha(\text{grad}[\varphi_{\ell^1}](\mathbf{a}), \mathcal{S}_\tau) + d_\alpha(\text{grad}[\varphi_\rho](\mathbf{a}) - \text{grad}[\varphi_{\ell^1}](\mathbf{a}), \mathcal{S}_\tau) \\ &\leq \|\boldsymbol{\beta}^* \chi_{\ell^1}[\boldsymbol{\beta}] \boldsymbol{\alpha}_{\tau^c} - \chi_{\ell^1}[\boldsymbol{\beta}]_{\tau^c}\|_2 + \|(\mathbf{I} - \boldsymbol{\alpha} \boldsymbol{\beta}^*)(\chi_\rho[\boldsymbol{\beta}] - \chi_{\ell^1}[\boldsymbol{\beta}])\|_2 \\ &\leq \frac{3}{2} n \theta \gamma + \frac{1}{20} n \theta \gamma + 4\sqrt{p} n \theta^4 \\ &\leq 3n \theta \gamma. \end{aligned} \quad (\text{J.31})$$

To bound the d_α norm of least eigenvector \mathbf{v} , note that $\boldsymbol{\beta}^* \chi_\rho[\boldsymbol{\beta}] > 0$, we can conclude

$$\mathbf{v}^* \nabla^2 \varphi_\rho(\mathbf{a}) \mathbf{v} \leq \mathbf{v}^* \mathbf{P}_{\mathbf{a}^\perp} \nabla^2 \varphi_\rho(\mathbf{a}) \mathbf{P}_{\mathbf{a}^\perp} \mathbf{v} + \boldsymbol{\beta}^* \chi_\rho[\boldsymbol{\beta}] = \mathbf{v}^* \mathbf{H} \mathbf{v} < -\eta_v,$$

expand $\nabla^2 \varphi_\rho(\mathbf{a})$ as in (H.8), and since \mathbf{v} is the eigenvector of smallest eigenvalue $\lambda_{\min} < -\eta_v$,

$$\mathbf{P}_{\mathbf{a}^\perp} \nabla^2 \varphi_\rho(\mathbf{a}) \mathbf{P}_{\mathbf{a}^\perp} \mathbf{v} = (\mathbf{I} - \mathbf{a} \mathbf{a}^*) \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \check{\mathbf{C}}_{x_0} \nabla_{\text{prox}_{\lambda \rho}}[\check{\mathbf{a}} * \mathbf{y}] \check{\mathbf{C}}_{x_0} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \mathbf{v} = \lambda_{\min} \mathbf{v}, \quad (\text{J.32})$$

hence there exists $\boldsymbol{\alpha}(\mathbf{v})$ satisfies $\mathbf{v} = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\alpha}(\mathbf{v})$ and

$$\boldsymbol{\alpha}(\mathbf{v}) = \lambda_{\min}^{-1} \left[\check{\mathbf{C}}_{x_0} \nabla_{\text{prox}_{\lambda \rho}}[\check{\mathbf{a}} * \mathbf{y}] \check{\mathbf{C}}_{x_0} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \mathbf{v} - \left(\boldsymbol{\beta}^* \check{\mathbf{C}}_{x_0} \nabla_{\text{prox}_{\lambda \rho}}[\check{\mathbf{a}} * \mathbf{y}] \check{\mathbf{C}}_{x_0} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} \mathbf{v} \right) \boldsymbol{\alpha} \right].$$

Now since $\nabla_{\text{prox}_{\lambda \rho}}[\check{\mathbf{a}} * \mathbf{y}]$ is a diagonal matrix with entries in $[0, 1]$,

$$d_\alpha(\mathbf{v}, \mathcal{S}_\tau) \leq \|\boldsymbol{\alpha}(\mathbf{v})\|_2 \leq |\lambda_{\min}|^{-1} \|\boldsymbol{\iota} \mathbf{C}_{\mathbf{a}_0}\|_2 \|\mathbf{x}_0\|_1^2 \|\mathbf{v}\|_2 (1 + \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2) < c_v n \theta p, \quad (\text{J.33})$$

where we use upper bound of $\|\mathbf{x}_0\|_1 < cn\theta$ from [Lemma D.2](#) and $|\lambda_{\min}| > \eta_v > cn\theta\lambda$ from [Corollary I.2](#).

3. (Iterates stay within widened subspace). Suppose [\(J.22\)](#) holds. We will show that whenever

$$t \leq T' = \frac{1}{10n\theta}, \quad (\text{J.34})$$

then setting $\mathbf{a}^+ = \mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}]$, we have

$$|d_\alpha(\mathbf{a}^+, \mathcal{S}_\tau) - d_\alpha(\mathbf{a}, \mathcal{S}_\tau)| \leq \frac{\gamma}{2}, \quad (\text{J.35})$$

and whenever $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) \in [\frac{\gamma}{2}, \gamma]$

$$d_\alpha^2(\mathbf{a}^+, \mathcal{S}_\tau) \leq d_\alpha^2(\mathbf{a}, \mathcal{S}_\tau) - t \cdot c'n\theta\gamma^2. \quad (\text{J.36})$$

If both [\(J.35\)](#) and [\(J.36\)](#) hold, then all iterates $\mathbf{a}^{(k)}$ will stay near the subspace: $d_\alpha(\mathbf{a}^{(k)}, \mathcal{S}_\tau) < \gamma$.

To derive [\(J.35\)](#), since both $\mathbf{g} \perp \mathbf{a}$ and $\mathbf{v} \perp \mathbf{a}$ we have $\|\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}\|_2^2 = \|\mathbf{a}\|_2^2 + \|t\mathbf{g} + t^2\mathbf{v}\|_2^2 > 1$, and since $d_\alpha(\cdot, \mathcal{S}_\tau)$ is a seminorm [Lemma E.2](#):

$$\begin{aligned} d_\alpha(\mathbf{a}^+, \mathcal{S}_\tau) &= d_\alpha(\mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}], \mathcal{S}_\tau) \leq d_\alpha(\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}, \mathcal{S}_\tau) \\ &\leq d_\alpha(\mathbf{a}, \mathcal{S}_\tau) + td_\alpha(\mathbf{g}, \mathcal{S}_\tau) + t^2d_\alpha(\mathbf{v}, \mathcal{S}_\tau) \end{aligned} \quad (\text{J.37})$$

suggests [\(J.35\)](#) holds via [\(J.28\)](#) and let $n > Cp^5\theta^{-2}$, we have

$$td_\alpha(\mathbf{g}, \mathcal{S}_\tau) + t^2d_\alpha(\mathbf{v}, \mathcal{S}_\tau) \leq \frac{c_g n\theta\gamma}{10n\theta} + \frac{c_v n\theta p}{(10n\theta)^2} < \frac{\gamma}{2} \quad (\text{J.38})$$

with sufficiently large C .

To derive [\(J.36\)](#), let $\boldsymbol{\alpha}(\mathbf{a})$ to be a coefficient vector satisfying $d_\alpha(\mathbf{a}, \mathcal{S}_\tau) = \|\boldsymbol{\alpha}(\mathbf{a})_{\tau^c}\|_2$, and based on [\(J.30\)](#) and [\(J.33\)](#), define

$$\boldsymbol{\alpha}(\mathbf{g}) = \boldsymbol{\beta}^* \boldsymbol{\chi}_\rho[\boldsymbol{\beta}] \boldsymbol{\alpha}(\mathbf{a}) - \boldsymbol{\chi}_\rho[\boldsymbol{\beta}] \quad (\text{J.39})$$

$$\boldsymbol{\alpha}(\mathbf{v}) = \lambda_{\min}^{-1} \check{\mathbf{C}}_{\mathbf{x}_0} \nabla_{\text{prox}_{\lambda\rho}}[\check{\mathbf{a}} * \mathbf{y}] \check{\mathbf{C}}_{\mathbf{x}_0} \mathbf{C}_{\mathbf{a}_0}^* \mathbf{v}. \quad (\text{J.40})$$

By the retraction property and norm bounds,

$$\langle \boldsymbol{\alpha}(\mathbf{a})_{\tau^c}, \boldsymbol{\alpha}(\mathbf{g})_{\tau^c} \rangle \geq \frac{1}{6n\theta} \|\boldsymbol{\alpha}(\mathbf{g})_{\tau^c}\|_2^2 \quad (\text{J.41})$$

$$\|\boldsymbol{\alpha}(\mathbf{a})_{\tau^c}\|_2 \leq \gamma \quad (\text{J.42})$$

$$\|\boldsymbol{\alpha}(\mathbf{v})\|_2 \leq c_v n\theta p. \quad (\text{J.43})$$

Since $\|\boldsymbol{\alpha}_{\tau^c}\|_2 > \frac{\gamma}{2}$,

$$\begin{aligned} \|\boldsymbol{\alpha}(\mathbf{g})_{\tau^c}\|_2 &\geq \|\boldsymbol{\beta}^* \boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}] \boldsymbol{\alpha}_{\tau^c} - \boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}]_{\tau^c}\|_2 - \|(\mathbf{I} - \boldsymbol{\alpha}\boldsymbol{\beta}^*)(\boldsymbol{\chi}_\rho[\boldsymbol{\beta}] - \boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}])\|_2 \\ &\geq |\boldsymbol{\beta}^* \boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}]| \|\boldsymbol{\alpha}_{\tau^c}\|_2 - \|\boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}]_{\tau^c}\|_2 - \|(\mathbf{I} - \boldsymbol{\alpha}\boldsymbol{\beta}^*)\|_2 \|(\boldsymbol{\chi}_\rho[\boldsymbol{\beta}] - \boldsymbol{\chi}_{\ell^1}[\boldsymbol{\beta}])\|_2 \\ &\geq \frac{1}{2}n\theta \times \frac{\gamma}{2} - \frac{1}{20}n\theta\gamma + 2n\theta^4 \\ &\geq \frac{1}{10}n\theta\gamma. \end{aligned} \quad (\text{J.44})$$

Finally, we can bound $d_\alpha(\mathbf{a}^+, \mathcal{S}_\tau)$ as

$$\begin{aligned} d_\alpha^2(\mathbf{a}^+, \mathcal{S}_\tau) &\leq d_\alpha^2(\mathbf{a} - t\mathbf{g} - t^2\mathbf{v}, \mathcal{S}_\tau) \\ &\leq \|[\boldsymbol{\alpha}(\mathbf{a}) - t\boldsymbol{\alpha}(\mathbf{g}) - t^2\boldsymbol{\alpha}(\mathbf{v})]_{\tau^c}\|_2^2 \\ &= \|\boldsymbol{\alpha}(\mathbf{a})_{\tau^c}\|_2^2 - 2t \langle \boldsymbol{\alpha}(\mathbf{a})_{\tau^c}, [\boldsymbol{\alpha}(\mathbf{g}) + t\boldsymbol{\alpha}(\mathbf{v})]_{\tau^c} \rangle + t^2 \|[\boldsymbol{\alpha}(\mathbf{g}) + t\boldsymbol{\alpha}(\mathbf{v})]_{\tau^c}\|_2^2 \end{aligned}$$

$$\begin{aligned}
&\leq \|\boldsymbol{\alpha}(\mathbf{a})_{\tau^c}\|_2^2 - 2t \langle \boldsymbol{\alpha}(\mathbf{a})_{\tau^c}, \boldsymbol{\alpha}(\mathbf{g})_{\tau^c} \rangle + 2t^2 \|\boldsymbol{\alpha}(\mathbf{a})_{\tau^c}\|_2 \|\boldsymbol{\alpha}(\mathbf{v})\|_2 \\
&\quad + 2t^2 \|\boldsymbol{\alpha}(\mathbf{g})_{\tau^c}\|_2^2 + 2t^4 \|\boldsymbol{\alpha}(\mathbf{v})\|_2^2 \\
&\leq d^2(\mathbf{a}, \mathcal{S}_\tau) - 2t \left[\left(\frac{1}{3n\theta} - t \right) \|\boldsymbol{\alpha}(\mathbf{g})_{\tau^c}\|_2^2 - tn\theta p\gamma - t^3(c_v n\theta p)^2 \right] \\
&\leq d^2(\mathbf{a}, \mathcal{S}_\tau) - t \cdot c' n\theta \gamma^2
\end{aligned} \tag{J.45}$$

where the last inequality holds when $t < \frac{0.1}{n\theta}$ with sufficiently large n .

4. (Polynomial time convergence) The iterates $\mathbf{a}^{(k)}$ remain within a γ neighborhood of \mathcal{S}_τ for all k . At any iteration k , $\mathbf{a}^{(k)}$ is in at least one of three regions: strong gradient, negative curvature, or strong convexity. In the gradient and curvature regions, we obtain a decrease in the function value which is at least some (nonzero) rational function of n and p . On the strongly convex region, the function value does not increase. The suboptimality at initialization is bounded by a polynomial in n and p , $\text{poly}(n, p)$, and hence the total number of steps in the gradient and curvature regions is bounded by a polynomial in n, p . After the iterates reach the strongly convex region, the number of additional steps required to achieve $\|\mathbf{a}^{(k)} - \bar{\mathbf{a}}\|_2 \leq \varepsilon$ is bounded by $\text{poly}(n, p) \log \varepsilon^{-1}$. In particular, the number of iterations required to achieve $\|\mathbf{a}^{(k)} - \bar{\mathbf{a}}\|_2 \leq \mu + 1/p$ is bounded by a polynomial in (n, p) , as claimed. ■

K Analysis of algorithm — local refinement

In this section, we describe and analyze an algorithm which refines an estimate $\mathbf{a}^{(0)} \approx \mathbf{a}_0$ of the kernel to exactly recover $(\mathbf{a}_0, \mathbf{x}_0)$. Set

$$\lambda^{(0)} \leftarrow 5\kappa_I \tilde{\mu} \quad \text{and} \quad I^{(0)} \leftarrow \text{supp}(\mathcal{S}_\lambda [\mathbf{C}_{\mathbf{a}^{(0)}}^* \mathbf{y}]), \tag{K.1}$$

where as each iteration of the algorithm consists of the following key steps:

- **Sparse Estimation using Reweighted Lasso:** Set

$$\mathbf{x}^{(k+1)} \leftarrow \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2} \|\mathbf{a}^{(k)} * \mathbf{x} - \mathbf{y}\|_2^2 + \sum_{i \notin I^{(k)}} \lambda^{(k)} |\mathbf{x}_i|; \tag{K.2}$$

- **Kernel Estimation using Least Squares:** Set

$$\mathbf{a}^{(k+1)} \leftarrow \mathbf{P}_{\mathbb{S}^{p-1}} \left[\underset{\mathbf{a}}{\text{argmin}} \frac{1}{2} \|\mathbf{a} * \mathbf{x}^{(k+1)} - \mathbf{y}\|_2^2 \right]; \tag{K.3}$$

- **Continuation and reweighting by decreasing sparsity regularizer:** Set

$$\lambda^{(k+1)} \leftarrow \frac{1}{2} \lambda^{(k)} \quad \text{and} \quad I^{(k+1)} \leftarrow \text{supp}(\mathbf{x}^{(k+1)}). \tag{K.4}$$

Our analysis will show that $\mathbf{a}^{(k)}$ converges to \mathbf{a}_0 at a linear rate. In the remainder of this section, we describe the assumptions of our analysis. In subsequent sections, we prove key lemmas analyzing each of the three main steps of the algorithm.

Below, we will write

$$\tilde{\mu} = \max \{ \mu, p^{-1} \}. \tag{K.5}$$

Our refinement algorithm will demand an initialization satisfying

$$\|\mathbf{a}^{(0)} - \mathbf{a}_0\|_2 \leq \tilde{\mu}. \tag{K.6}$$

Our goal is to show that the proposed annealing algorithm exactly solves the SaS deconvolution problem, i.e., exactly recovers $(\mathbf{a}_0, \mathbf{x}_0)$ up to a signed shift. We will denote the support sets of true sparse vector \mathbf{x}_0 and recovered $\mathbf{x}^{(k)}$ in the intermediate k -th steps as

$$I = \text{supp}(\mathbf{x}_0), \quad I^{(k)} = \text{supp}(\mathbf{x}^{(k)}). \quad (\text{K.7})$$

It should be clear that exact recovery is unlikely if \mathbf{x}_0 contains many consecutive nonzero entries: in this situation, even *non-blind* deconvolution fails. We introduce the notation κ_I as an upper bound for number of nonzero entries of \mathbf{x}_0 in a length- p window:

$$\kappa_I = 6 \max\{\theta p, \log n\}, \quad (\text{K.8})$$

then in the Bernoulli-Gaussian model, with high probability,

$$\max_{\ell} |I \cap ([p] + \ell)| \leq \kappa_I. \quad (\text{K.9})$$

Here, indexing and addition should be interpreted modulo n . The $\log n$ term reflects the fact that as n becomes enormous (exponential in p) eventually it becomes likely that some length- p window of \mathbf{x}_0 is densely occupied. In our main theorem statement, we preclude this possibility by putting an upper bound on n (w.r.t $\tilde{\mu}$). We find it useful to also track the maximum ℓ^2 norm of \mathbf{x}_0 over any length- p window:

$$\|\mathbf{x}_0\|_{\square} := \max_{\ell} \|\mathbf{P}_{([p] + \ell)} \mathbf{x}_0\|_2. \quad (\text{K.10})$$

Below, we will sometimes work with the \square -induced operator norm:

$$\|\mathbf{M}\|_{\square \rightarrow \square} = \sup_{\|\mathbf{x}\|_{\square} \leq 1} \|\mathbf{M}\mathbf{x}\|_{\square} \quad (\text{K.11})$$

For now, we note that in the Bernoulli-Gaussian model, $\|\mathbf{x}_0\|_{\square}$ is typically not large

$$\|\mathbf{x}_0\|_{\square} \leq \sqrt{\kappa_I}. \quad (\text{K.12})$$

K.1 Reweighted Lasso finds the large entries of \mathbf{x}_0

The following lemma asserts that when \mathbf{a} is close to \mathbf{a}_0 , the reweighted Lasso finds all of the large entries of \mathbf{x}_0 . Our reweighted Lasso is modified version from [CWB08], we only penalize \mathbf{x} on entries outside of its known support subset. We write T to be the subset of true support I , and define the sparsity surrogate as

$$\sum_{i \in T^c} |\mathbf{x}_i| \quad (\text{K.13})$$

The reweighted Lasso recovers more accurate \mathbf{x} on set T compares to the vanilla Lasso problem, it turns out to be very helpful in our analysis which proves convergence of the proposed alternating minimization.

Lemma K.1 (Accuracy of reweighted Lasso estimate). *Suppose that $\mathbf{y} = \mathbf{a}_0 * \mathbf{x}_0$ with \mathbf{a}_0 is $\tilde{\mu}$ -shift coherent and $\|\mathbf{x}_0\|_{\square} \leq \sqrt{\kappa_I}$ with $\kappa_I \geq 1$. If $\tilde{\mu}\kappa_I^2 \leq c_{\mu}$, then for every $T \subseteq I$ and \mathbf{a} satisfying $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$, the solution \mathbf{x}^+ to the optimization problem*

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{a} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{i \in T^c} |\mathbf{x}_i| \right\}, \quad (\text{K.14})$$

with

$$\lambda > 5\kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2, \quad (\text{K.15})$$

is unique with the form

$$\mathbf{x}^+ = \boldsymbol{\iota}_J (\mathbf{C}_{\mathbf{a}, J}^* \mathbf{C}_{\mathbf{a}, J})^{-1} \boldsymbol{\iota}_J^* (\mathbf{C}_{\mathbf{a}}^* \mathbf{y} - \lambda \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}) \quad (\text{K.16})$$

where $\boldsymbol{\sigma} = \text{sign}(\mathbf{x}^+)$. Its support set J satisfies

$$(T \cup I_{\geq 3\lambda}) \subseteq J \subseteq I \quad (\text{K.17})$$

and its entrywise error is bounded as

$$\|\mathbf{x}^+ - \mathbf{x}_0\|_\infty \leq 3\lambda. \quad (\text{K.18})$$

Above, $c_\mu > 0$ is a positive numerical constant.

We prove [Lemma K.1](#) below. The proof relies heavily on the fact that when \mathbf{a}_0 is shift-incoherent and $\mathbf{a} \approx \mathbf{a}_0$, \mathbf{a} is also shift-incoherent, an observation which is formalized in a sequence of calculations in [Appendix K.4](#).

Proof. 1. (Restricted support Lasso problem). We first consider the restricted problem

$$\min_{\mathbf{w} \in \mathbb{R}^{|I|}} \left\{ \frac{1}{2} \|\mathbf{a} * \boldsymbol{\iota}_I \mathbf{w} - \mathbf{y}\|_2^2 + \lambda \sum_{i \in T^c} |(\boldsymbol{\iota}_I \mathbf{w})_i| \right\}. \quad (\text{K.19})$$

Under our assumptions, provided $c < \frac{1}{9}$, [Lemma K.6](#) implies that

$$\boldsymbol{\iota}_I^* \mathbf{C}_a^* \mathbf{C}_a \boldsymbol{\iota}_I = [\mathbf{C}_a^* \mathbf{C}_a]_{I,I} \succ \mathbf{0}, \quad (\text{K.20})$$

and the restricted problem is strongly convex and its solution is unique. The KKT conditions imply that a vector \mathbf{w}_* is the unique optimal solution to this problem if and only if

$$\boldsymbol{\iota}_I^* \mathbf{C}_a^* \mathbf{C}_a \boldsymbol{\iota}_I \mathbf{w}_* \in \boldsymbol{\iota}_I^* \mathbf{C}_a^* \mathbf{y} - \lambda \partial \|\mathbf{P}_{T^c}[\cdot]\|_1(\mathbf{w}_*). \quad (\text{K.21})$$

Writing $J = \text{supp}(\boldsymbol{\iota}_I \mathbf{w}_*) \subseteq I$, $\mathbf{C}_{a,J} = \mathbf{C}_a \boldsymbol{\iota}_J$, $\mathbf{w}_J = \boldsymbol{\iota}_J^* \boldsymbol{\iota}_I \mathbf{w}_*$ the corresponding sub-vector containing the nonzero entries of \mathbf{w}_* and $\boldsymbol{\sigma}_{J \setminus T} = \boldsymbol{\iota}_J^* \mathbf{P}_{T^c}[\text{sign}(\boldsymbol{\iota}_I \mathbf{w}_*)]$, the condition [\(K.21\)](#) is satisfied if and only if

$$\mathbf{C}_{a,J}^* \mathbf{C}_{a,J} \mathbf{w}_J = \mathbf{C}_{a,J}^* \mathbf{y} - \lambda \boldsymbol{\sigma}_{J \setminus T}, \quad (\text{K.22})$$

$$\|\mathbf{C}_{a_{I \setminus J}}^* (\mathbf{C}_{a,J} \mathbf{w}_J - \mathbf{y})\|_\infty \leq \lambda. \quad (\text{K.23})$$

We will argue that under our assumptions, J necessarily contains all of the large entries of \mathbf{x}_0 :

$$I_{>3\lambda} = \{\ell \in I \mid |\mathbf{x}_{0\ell}| > 3\lambda\} \subseteq J. \quad (\text{K.24})$$

We show this by contradiction – namely, if some large entry of \mathbf{x}_0 is not in J , then the dual condition [\(K.23\)](#) is violated, contradicting the optimality of \mathbf{w}_* . To this end, note that by [Corollary K.7](#), $\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}$ has full rank. From [\(K.22\)](#),

$$\mathbf{w}_J = [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} [\mathbf{C}_{a,J}^* \mathbf{y} - \lambda \boldsymbol{\sigma}_{J \setminus T}]. \quad (\text{K.25})$$

Write $\mathbf{x}_{0,J} = \boldsymbol{\iota}_J^* \mathbf{x}_0$ and $(\mathbf{x}_0)_{I \setminus J} = \mathbf{P}_{I \setminus J} \mathbf{x}_0$. We can further notice that

$$\begin{aligned} \mathbf{C}_{a,J} \mathbf{w}_J - \mathbf{y} &= \left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{y} - \lambda \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \boldsymbol{\sigma}_{J \setminus T} \\ &= \left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{C}_{a_0,J} \mathbf{x}_{0,J} \\ &\quad + \left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{C}_{a_0, I \setminus J} (\mathbf{x}_0)_{I \setminus J} \\ &\quad - \lambda \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \boldsymbol{\sigma}_{J \setminus T} \\ &= \left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{C}_{a_0 - a, J} \mathbf{x}_{0,J} \\ &\quad + \left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{C}_{a_0, I \setminus J} (\mathbf{x}_0)_{I \setminus J} \\ &\quad - \lambda \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \boldsymbol{\sigma}_{J \setminus T}, \end{aligned} \quad (\text{K.26})$$

where in the final line we have used that

$$\left(\mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* - \mathbf{I} \right) \mathbf{C}_{a,J} = \mathbf{0}. \quad (\text{K.27})$$

Suppose that J is a strict subset of I (otherwise, if $J = I$, we are done). Take any $i \in I \setminus J$ such that $|\mathbf{x}_{0i}| = \|(\mathbf{x}_0)_{I \setminus J}\|_\infty$, and let $\xi = \text{sign}(\mathbf{x}_{0i})$. Using (K.26), Corollary K.7 and Lemma K.8, and simplify the induced norms $\|\cdot\|_{\infty \rightarrow \infty}$ and $\|\cdot\|_{\square \rightarrow \square}$ as $\|\cdot\|_\infty$ and $\|\cdot\|_\square$, we have

$$\begin{aligned} -\xi s_i[\mathbf{a}]^* (\mathbf{C}_{a,J} \mathbf{w}_J - \mathbf{y}) &= \xi s_i[\mathbf{a}]^* \left(\mathbf{I} - \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* \right) s_i[\mathbf{a}_0] \mathbf{x}_{0i} \\ &\quad + \xi s_i[\mathbf{a}]^* \left(\mathbf{I} - \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* \right) \mathbf{C}_{a_0} (\mathbf{x}_0)_{I \setminus (J \cup \{i\})} \\ &\quad + \xi s_i[\mathbf{a}]^* \left(\mathbf{I} - \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \mathbf{C}_{a,J}^* \right) \mathbf{C}_{a_0 - a, J} \mathbf{x}_{0J} \\ &\quad + \xi \lambda s_i[\mathbf{a}]^* \mathbf{C}_{a,J} [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \boldsymbol{\sigma}_{J \setminus T} \end{aligned} \quad (\text{K.28})$$

$$\begin{aligned} &\geq \left(\langle s_i[\mathbf{a}], s_i[\mathbf{a}_0] \rangle \right. \\ &\quad - \|s_i[\mathbf{a}]^* \mathbf{C}_{a,J}\|_1 \left\| [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \right\|_\infty \| \mathbf{C}_{a,J}^* s_i[\mathbf{a}_0] \|_\infty \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad - \left(\left\| s_i[\mathbf{a}]^* \mathbf{C}_{a_0 I \setminus \{i\}} \right\|_1 \right. \\ &\quad \left. + \|s_i[\mathbf{a}]^* \mathbf{C}_{a,J}\|_1 \left\| [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \right\|_\infty \left\| \mathbf{C}_{a,J}^* \mathbf{C}_{a_0 I \setminus J} \right\|_\infty \right) \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad - \left(\|s_i[\mathbf{a}]^* \mathbf{C}_{a_0 - a, J}\|_2 \right. \\ &\quad \left. + \|s_i[\mathbf{a}]^* \mathbf{C}_{a,J}\|_2 \left\| [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \right\|_\square \left\| \mathbf{C}_{a,J}^* \mathbf{C}_{a_0 - a, J} \right\|_\square \right) \sqrt{2} \|\mathbf{x}_0\|_\square \\ &\quad - \lambda \|s_i[\mathbf{a}]^* \mathbf{C}_{a,J}\|_1 \left\| [\mathbf{C}_{a,J}^* \mathbf{C}_{a,J}]^{-1} \right\|_\infty \|\boldsymbol{\sigma}_{J \setminus T}\|_\infty \end{aligned} \quad (\text{K.29})$$

$$\begin{aligned} &\geq \left((1 - \|\mathbf{a} - \mathbf{a}_0\|_2) - C_1 \kappa_I \tilde{\mu} \times 1 \times \tilde{\mu} \right) \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad - C_2 \left(\kappa_I \tilde{\mu} + \kappa_I \tilde{\mu} \times 1 \times \kappa_I \tilde{\mu} \right) \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad - \left(2\sqrt{\kappa_I} \|\mathbf{a} - \mathbf{a}_0\|_2 + C_3 \sqrt{\kappa_I} \tilde{\mu} \times 1 \times \kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2 \right) \|\mathbf{x}_0\|_\square \\ &\quad - \lambda C_4 \kappa_I \tilde{\mu} \end{aligned} \quad (\text{K.30})$$

$$\begin{aligned} &\geq \left(1 - C'_1 \kappa_I \tilde{\mu} - C_2 (\kappa_I \tilde{\mu})^2 \right) \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad - 2\kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2 - \left(C_3 \kappa_I^{3/2} \tilde{\mu} \right) \kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2 - (C_4 \kappa_I \tilde{\mu}) \lambda \end{aligned} \quad (\text{K.31})$$

$$\geq \frac{1}{2} \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty - \lambda/2, \quad (\text{K.32})$$

where the last line holds provided $\tilde{\mu} \kappa_I^2 \leq c_\mu$ to be a sufficiently small numerical constants. If $\left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty > 3\lambda$, this is strictly larger than λ , implying that $|\mathbf{a}_i^* (\mathbf{C}_{a,J} \mathbf{w}_J - \mathbf{y})| > \lambda$, and contradicting the KKT conditions for the restricted problem. Hence, under our assumptions

$$\left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \leq 3\lambda. \quad (\text{K.33})$$

2. (Solution of Full Lasso problem) We next argue that the solution of the restricted support Lasso problem, \mathbf{w}_J , when extended to \mathbb{R}^n as $\mathbf{x}^+ = \boldsymbol{\iota}_J \mathbf{w}_J$, is the unique optimal solution to the *full* Lasso problem

$$\min_{\mathbf{x}} \varphi_{\text{lasso}}(\mathbf{x}) \equiv \frac{1}{2} \|\mathbf{a} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{i \in T^c} |\mathbf{x}_i|. \quad (\text{K.34})$$

To prove that \mathbf{x}^+ is the unique optimal solution, it suffices to show that for every $i \in I^c$,

$$|s_i[\mathbf{a}]^*(\mathbf{a} * \mathbf{x}^+ - \mathbf{y})| < \lambda. \quad (\text{K.35})$$

Indeed, suppose that this inequality is in force. Write $\varepsilon = \lambda - \max_{i \in I^c} |s_i[\mathbf{a}]^*(\mathbf{a} * \mathbf{x}^+ - \mathbf{y})|$, and notice that from the KKT conditions for the restricted problem,

$$\mathbf{0} \in \mathbf{P}_I \partial_{\mathbf{x}} \varphi_{\text{lasso}}(\mathbf{x}) \quad (\text{K.36})$$

Combining with (K.35), we have that for every vector ζ with $\text{supp}(\zeta) \subseteq I^c$ and $\|\zeta\|_\infty \leq 1$, then $\varepsilon \zeta \in \partial \varphi_{\text{lasso}}(\mathbf{x}^+)$. Let \mathbf{x}' be any vector with $\mathbf{x}'_{I^c} \neq \mathbf{0}$ and set $\zeta = \mathcal{P}_{I^c} \text{sign}(\mathbf{x}')$, then from the subgradient inequality,

$$\begin{aligned} \varphi_{\text{lasso}}(\mathbf{x}') &\geq \varphi_{\text{lasso}}(\mathbf{x}^+) + \langle \varepsilon \zeta, \mathbf{x}' - \mathbf{x}^+ \rangle \\ &\geq \varphi_{\text{lasso}}(\mathbf{x}^+) + \varepsilon \|\mathbf{x}'_{I^c}\|_1, \end{aligned} \quad (\text{K.37})$$

which is strictly larger than $\varphi_{\text{lasso}}(\mathbf{x}^+)$. Hence, when (K.35) holds, any optimal solution $\bar{\mathbf{x}}$ to the full Lasso problem must satisfy $\text{supp}(\bar{\mathbf{x}}) \subseteq I$. By strong convexity of the restricted problem, the solution to (K.34) is unique and equal to \mathbf{x}^+ .

We finish by showing (K.35). Using the same expansion as above, we obtain

$$\begin{aligned} |s_i[\mathbf{a}]^*(\mathbf{C}_{aJ} \mathbf{w}_J - \mathbf{y})| &\leq \left| s_i[\mathbf{a}]^* \left(\mathbf{I} - \mathbf{C}_{aJ} [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \mathbf{C}_{aJ}^* \right) \mathbf{C}_{a_0 I \setminus J}(\mathbf{x}_0)_{I \setminus J} \right| \\ &\quad + \left| s_i[\mathbf{a}]^* \left(\mathbf{I} - \mathbf{C}_{aJ} [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \mathbf{C}_{aJ}^* \right) \mathbf{C}_{a_0 - aJ} \mathbf{x}_{0J} \right| \\ &\quad + \lambda \left| s_i[\mathbf{a}]^* \mathbf{C}_{aJ} [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \boldsymbol{\sigma}_{J \setminus T} \right| \end{aligned} \quad (\text{K.38})$$

$$\begin{aligned} &\leq \left(\left\| s_i[\mathbf{a}]^* \mathbf{C}_{a_0 I \setminus J} \right\|_1 \right. \\ &\quad \left. + \left\| s_i[\mathbf{a}]^* \mathbf{C}_{aJ} \right\|_1 \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \right\|_\infty \left\| \mathbf{C}_{aJ}^* \mathbf{C}_{a_0 I \setminus J} \right\|_\infty \right) \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\quad + \left(\left\| s_i[\mathbf{a}]^* \mathbf{C}_{a_0 - aJ} \right\|_2 \right. \\ &\quad \left. + \left\| s_i[\mathbf{a}]^* \mathbf{C}_{aJ} \right\|_2 \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \right\|_\square \left\| \mathbf{C}_{aJ}^* \mathbf{C}_{a_0 - aJ} \right\|_\square \right) \sqrt{2} \left\| \mathbf{x}_0 \right\|_\square \\ &\quad + \lambda \left\| s_i[\mathbf{a}]^* \mathbf{C}_{aJ} \right\|_1 \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \right\|_\infty \left\| \boldsymbol{\sigma}_{J \setminus T} \right\|_\infty \end{aligned} \quad (\text{K.39})$$

$$\begin{aligned} &\leq C_1 (\tilde{\mu} \kappa_I + \tilde{\mu} \kappa_I \times 1 \times \tilde{\mu} \kappa_I) \times 2\lambda \\ &\quad + (2\sqrt{\kappa_I} \|\mathbf{a} - \mathbf{a}_0\|_2 + C_2 \sqrt{\kappa_I} \tilde{\mu} \times 1 \times \kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2) \times \sqrt{\kappa_I} \\ &\quad + \lambda C_3 \times \tilde{\mu} \kappa_I \end{aligned} \quad (\text{K.40})$$

$$\leq ((C_1 + C_3) \tilde{\mu} \kappa_I + C_1 (\tilde{\mu} \kappa_I)^2) \lambda + (2 + C_2 \tilde{\mu} \kappa_I) \kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2 \quad (\text{K.41})$$

$$< \lambda, \quad (\text{K.42})$$

where the last line holds as long as c_μ is a sufficiently small numerical constant. This establishes that \mathbf{x}^+ is the unique optimal solution to the full Lasso problem.

3. (Entrywise difference to \mathbf{x}_0) Finally we will be controlling $\|\mathbf{x}_J^+ - (\mathbf{x}_0)_J\|_\infty$. Indeed, from [Corollary K.7](#), [Lemma K.8](#),

$$\begin{aligned} \|\mathbf{x}_J^+ - (\mathbf{x}_0)_J\|_\infty &= \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \mathbf{C}_{aJ}^* \mathbf{C}_{a_0} \mathbf{x}_0 - \lambda [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \boldsymbol{\sigma}_{J \setminus T} - (\mathbf{x}_0)_J \right\|_\infty \\ &\leq \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \mathbf{C}_{aJ}^* \mathbf{C}_{a_0 - aJ} (\mathbf{x}_0)_J \right\|_\infty + \lambda \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \boldsymbol{\sigma}_{J \setminus T} \right\|_\infty \\ &\quad + \left\| [\mathbf{C}_{aJ}^* \mathbf{C}_{aJ}]^{-1} \mathbf{C}_{aJ}^* \mathbf{C}_{aI \setminus J} (\mathbf{x}_0)_{I \setminus J} \right\|_\infty \\ &\leq 2 \left\| \mathbf{C}_{aJ}^* \mathbf{C}_{a_0 - aJ} \right\|_{\square \rightarrow \infty} \left\| (\mathbf{x}_0)_J \right\|_\square + \end{aligned}$$

$$\begin{aligned}
& 2\lambda + 2 \left\| \mathbf{C}_{\mathbf{a}^* J} \mathbf{C}_{\mathbf{a}^* I \setminus J} \right\|_{\infty} \left\| (\mathbf{x}_0)_{I \setminus J} \right\|_{\infty} \\
& \leq 2\sqrt{2\kappa_I} \|\mathbf{a} - \mathbf{a}_0\|_2 \|\mathbf{x}_0\|_{\square} + 2\lambda + 2 \times 3\tilde{\mu} \times 2\kappa_{I \setminus J} \times 3\lambda \\
& \leq 3\kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2 + 2\lambda + 36\lambda\tilde{\mu}\kappa_I \\
& \leq 3\lambda,
\end{aligned} \tag{K.43}$$

establishing the claim. ■

K.2 Least squares solution $\mathbf{a}^{(k)}$ contracts

In this section, given \mathbf{x} to be the solution to the reweighted Lasso from \mathbf{a} , we will show the solution of the least squares problem

$$\mathbf{a}^+ \leftarrow \operatorname{argmin}_{\mathbf{a}' \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{a}' * \mathbf{x} - \mathbf{y}\|_2^2 \tag{K.44}$$

is closer to \mathbf{a}_0 compared to \mathbf{a} . Observe that in [Lemma K.1](#), the solution of [\(K.16\)](#)

$$\mathbf{x} = \iota_J (\mathbf{C}_{\mathbf{a}^* J}^* \mathbf{C}_{\mathbf{a}^* J})^{-1} \iota_J^* (\mathbf{C}_{\mathbf{a}^*}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0 - \lambda \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}), \tag{K.45}$$

by assuming $\mathbf{C}_{\mathbf{a}^* J}^* \mathbf{C}_{\mathbf{a}^* J} \approx \mathbf{I}$, $\mathbf{a} \approx \mathbf{a}_0$ and $J \setminus T \approx \emptyset$, is a good approximation to the true sparse map \mathbf{x}_0

$$\mathbf{x} \approx \mathbf{I}(\mathbf{x}_0 - \mathbf{0}) = \mathbf{x}_0; \tag{K.46}$$

furthermore, its difference to the true sparse map $\|\mathbf{x}_0 - \mathbf{x}\|_2$ is proportional to $\|\mathbf{a}_0 - \mathbf{a}\|_2$ as

$$\mathbf{x} - \mathbf{x}_0 \approx \mathbf{P}_I (\mathbf{C}_{\mathbf{a}^*}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0 - \mathbf{C}_{\mathbf{a}^*}^* \mathbf{C}_{\mathbf{a}} \mathbf{x}_0) \approx \mathbf{P}_I [\mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \iota(\mathbf{a}_0 - \mathbf{a})]. \tag{K.47}$$

To this end, since we know the solution of least square problem \mathbf{a}^+ is simply

$$\mathbf{a}^+ = (\iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}} \iota)^{-1} (\iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}_0} \iota \mathbf{a}_0), \tag{K.48}$$

this implies the difference between the new \mathbf{a}^+ and \mathbf{a}_0 , has the relationship with $\mathbf{a} - \mathbf{a}_0$ roughly

$$\begin{aligned}
\mathbf{a}^+ - \mathbf{a}_0 &= (\iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}} \iota)^{-1} (\iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}_0} \iota \mathbf{a}_0 - \iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}} \iota \mathbf{a}_0) \\
&\approx (n\theta)^{-1} \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} (\mathbf{x}_0 - \mathbf{x}) \\
&\approx (n\theta)^{-1} \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_I \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \iota (\mathbf{a} - \mathbf{a}_0).
\end{aligned} \tag{K.49}$$

To make this point precise, we introduce the following lemma:

Lemma K.2 (Approximation of least square estimate). *Given $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ to be $\tilde{\mu}$ -shift coherent and $\mathbf{x}_0 \sim \text{BG}(\theta) \in \mathbb{R}^n$. There exists some constants C, C', c, c', c_{μ} such that if $\lambda < c' \tilde{\mu} \kappa_I$, $\tilde{\mu} \kappa_I^2 \leq c_{\mu}$ and $n > Cp^2 \log p$, then with probability at least $1 - c/n$, for every \mathbf{a} satisfying $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$ and \mathbf{x} of the form*

$$\mathbf{x} = \iota_J (\mathbf{C}_{\mathbf{a}^* J}^* \mathbf{C}_{\mathbf{a}^* J})^{-1} \iota_J^* (\mathbf{C}_{\mathbf{a}^*}^* \mathbf{y} - \lambda \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}) \tag{K.50}$$

where the set J, T satisfies $I_{>6\lambda} \subseteq T \subseteq J \subseteq I$, we have

$$\begin{aligned}
& \frac{1}{n\theta} \left\| \iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x} - \mathbf{x}_0} \iota \mathbf{a}_0 - \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_I \mathbf{C}_{\mathbf{a}_0}^* \mathbf{C}_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \right\|_2 \\
& \leq C' \lambda \left(\tilde{\lambda} + \tilde{\mu} \kappa_I \right) + \frac{1}{32} \|\mathbf{a} - \mathbf{a}_0\|_2
\end{aligned} \tag{K.51}$$

with $\tilde{\lambda} = \lambda + \frac{\log n}{\sqrt{n\theta^2}}$.

Proof. We will begin with listing the conditions we use for both \mathbf{x} and \mathbf{x}_0 . First, we know from [Lemma K.1](#) and our assumptions on the set T , then \mathbf{x} approximates \mathbf{x}_0 in the sense that

$$\|\mathbf{x} - \mathbf{x}_0\|_\infty \leq 3\lambda \quad (\text{K.52})$$

$$\|(\mathbf{x}_0)_{I \setminus J}\|_\infty \leq 3\lambda \quad (\text{K.53})$$

$$\|(\mathbf{x}_0)_{I \setminus T}\|_\infty \leq 6\lambda. \quad (\text{K.54})$$

Write $\mathbf{x}_0 = \mathbf{g} \circ \boldsymbol{\omega}$ with \mathbf{g} iid standard normal, $\boldsymbol{\omega}$ iid Bernoulli and \mathbf{g} and $\boldsymbol{\omega}$ independent. From [\(K.53\)](#) we know $|I \setminus J| = |\{i \mid |\mathbf{g}_i| \leq 3\lambda, \boldsymbol{\omega}_i \neq 0\}|$. Since $\mathbb{P}[\boldsymbol{\omega}_i \neq 0] = \theta$ and $\mathbb{P}[|\mathbf{g}_i| \leq 3\lambda] \leq 3\lambda$, [Lemma D.1](#) implies that with probability at least $1 - 2/n$:

$$|I \setminus J| \leq 3\lambda n\theta + 6\sqrt{\lambda n\theta} \log n \leq 3\tilde{\lambda}n\theta \quad (\text{K.55})$$

$$|I \setminus T| \leq 6\lambda n\theta + 12\sqrt{\lambda n\theta} \log n \leq 6\tilde{\lambda}n\theta, \quad (\text{K.56})$$

and

$$|(I \setminus J) \cap s_\ell[I]| \leq 3\lambda n\theta^2 + 6\sqrt{\lambda n\theta^2} \log n \leq 3\tilde{\lambda}n\theta^2; \quad (\text{K.57})$$

together with base on properties of Bernoulli-Gaussian vector \mathbf{x}_0 from [Appendix D](#) and we conclude with probability at least $1 - c/n$, all the following events hold:

$$\frac{1}{2}n\theta \leq |I| \leq 2n\theta, \quad (\text{K.58})$$

$$\max_{\ell \neq 0} |I \cap s_\ell[I]| \leq 2n\theta^2 \quad (\text{K.59})$$

$$\max_{\ell \neq 0} |(I \setminus J) \cap s_\ell[I]| \leq 6\tilde{\lambda}n\theta^2, \quad (\text{K.60})$$

$$\|\mathbf{x}_0\|_\square^2 \leq \kappa_I, \quad (\text{K.61})$$

$$\|\check{\mathbf{a}}_0 * \mathbf{x}_0\|_\square^2 \leq \kappa_I, \quad (\text{K.62})$$

$$\|\mathbf{x}_0\|_2^2 \leq 2n\theta, \quad (\text{K.63})$$

$$\|\mathbf{x}_0\|_1 \leq 2n\theta, \quad (\text{K.64})$$

$$\max_{\ell \neq 0} \|\mathbf{P}_{I \cap s_\ell[I]} \mathbf{x}_0\|_2^2 \leq 2n\theta^2, \quad (\text{K.65})$$

$$\max_{\ell \neq 0} \|\mathbf{P}_{I \cap s_\ell[I \setminus J]} \mathbf{x}_0\|_1 \leq 12\tilde{\lambda}n\theta^2, \quad (\text{K.66})$$

$$\|\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}\|_2^2 \leq 3n\theta, \quad (\text{K.67})$$

provided by $n \geq C\theta^{-2} \log p$ for sufficiently large constant C .

1. (Approximate $\mathbf{C}_{\mathbf{x}}$ with $\mathbf{C}_{\mathbf{x}_0}$) Since

$$\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 = \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 + \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 \quad (\text{K.68})$$

where

$$\begin{aligned} \|\boldsymbol{\iota}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0\|_2 &\leq \|\mathbf{a}_0\|_2 \|\mathbf{x} - \mathbf{x}_0\|_2^2 + \|\mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_2 \sqrt{2p} \max_{\ell \neq 0} |\langle s_\ell[\mathbf{x} - \mathbf{x}_0], \mathbf{x} - \mathbf{x}_0 \rangle| \\ &\leq \|\mathbf{x} - \mathbf{x}_0\|_\infty^2 \times |I| + \sqrt{2\tilde{\mu}p^2} \left(\|\mathbf{x} - \mathbf{x}_0\|_\infty^2 \times \max_{\ell \neq 0} |I \cap s_\ell[I]| \right) \\ &\leq C_1 \left(\lambda^2 n\theta + \sqrt{2\tilde{\mu}p^2} (\lambda^2 n\theta^2) \right) \\ &\leq 2C_1 \lambda^2 n\theta, \end{aligned} \quad (\text{K.69})$$

we have that

$$\|\iota^* \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 - \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0\|_2 \leq 2C_1 \lambda^2 n \theta. \quad (\text{K.70})$$

2. (Extract the $\mathbf{a}_0 - \mathbf{a}$ term) Observe that

$$\begin{aligned} \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota} \mathbf{a}_0 &= \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} (\mathbf{x} - \mathbf{x}_0) \\ &= \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \left(\boldsymbol{\iota}_J (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \boldsymbol{\iota}_J^* (\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0 - \lambda \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}) \right. \\ &\quad \left. - \boldsymbol{\iota}_J (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J}) (\mathbf{x}_0)_J - \mathbf{P}_{I \setminus J} \mathbf{x}_0 \right) \\ &= \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0 J} (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \mathbf{C}_{\mathbf{a}_J}^* (\mathbf{C}_{\mathbf{a}_0 - \mathbf{a}} \mathbf{x}_0) \\ &\quad + \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0 J} (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \mathbf{C}_{\mathbf{a}_J}^* (\mathbf{C}_{\mathbf{a}} \mathbf{x}_0 - \mathbf{C}_{\mathbf{a}_J} (\mathbf{x}_0)_J) \\ &\quad - \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_{I \setminus J} \mathbf{x}_0 \\ &\quad - \lambda \iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0 J} (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \boldsymbol{\iota}_J^* \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}, \end{aligned} \quad (\text{K.71})$$

where, the second term in (K.71) is bounded as

$$\begin{aligned} &\|\iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0 J} (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \mathbf{C}_{\mathbf{a}_J}^* (\mathbf{C}_{\mathbf{a}} \mathbf{x}_0 - \mathbf{C}_{\mathbf{a}_J} (\mathbf{x}_0)_J)\|_2 \\ &\leq \|\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}\|_2 \times \|\mathbf{C}_{\mathbf{a}_0 J}\|_2 \|\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J}\|_2^{-1} \\ &\quad \times \|\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J}\|_2 \times \|(\mathbf{x}_0)_{I \setminus J}\|_2 \\ &\leq C_2 \left(\sqrt{n\theta} \times 3 \times \tilde{\mu} \kappa_I \times \lambda \sqrt{\tilde{\lambda} n \theta} \right) \\ &\leq 3C_2 \tilde{\mu} \kappa_I \lambda n \theta; \end{aligned} \quad (\text{K.72})$$

the third term in (K.71) is bounded as

$$\begin{aligned} \|\iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{P}_{I \setminus J} \mathbf{x}_0\|_2 &= \|\iota^* \mathbf{C}_{\mathbf{a}_0} (\mathbf{P}_{[\pm p] \setminus 0} + \mathbf{e}_0 \mathbf{e}_0^*) \mathbf{C}_{\mathbf{x}_0}^* \mathbf{P}_{I \setminus J} \mathbf{x}_0\|_2 \\ &\leq \|\mathbf{a}_0\|_2 \|(\mathbf{x}_0)_{I \setminus J}\|_2^2 \\ &\quad + \|\mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_2 \times \sqrt{2p} \times \max_{\ell \neq 0} \|\mathbf{P}_{I \cap s_\ell [I \setminus J]} \mathbf{x}_0\|_1 \times \|(\mathbf{x}_0)_{I \setminus J}\|_\infty \\ &\leq C_3 \left(\lambda^2 \times \tilde{\lambda} n \theta + \sqrt{\tilde{\mu} p^2} \times \tilde{\lambda} n \theta^2 \times \lambda \right) \\ &\leq 2C_3 \tilde{\lambda} \lambda n \theta; \end{aligned} \quad (\text{K.73})$$

and finally, write $\boldsymbol{\Delta} = (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} - \mathbf{I}$, then the fourth term in (K.71) is bounded as

$$\begin{aligned} &\lambda \|\iota^* \mathbf{C}_{\mathbf{x}_0}^* \mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}_J (\mathbf{C}_{\mathbf{a}_J}^* \mathbf{C}_{\mathbf{a}_J})^{-1} \boldsymbol{\iota}_J^* \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}\|_2 \\ &= \lambda \|\iota^* \mathbf{C}_{\mathbf{a}_0} (\mathbf{P}_{[\pm p] \setminus 0} + \mathbf{e}_0 \mathbf{e}_0^*) \mathbf{C}_{\mathbf{x}_0}^* \boldsymbol{\iota}_J (\mathbf{I} + \boldsymbol{\Delta}) \boldsymbol{\iota}_J^* \mathbf{P}_{J \setminus T} \boldsymbol{\sigma}\|_2 \\ &\leq \lambda \|\mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_2 \sqrt{2p} \max_{\ell \neq 0} \|\mathbf{P}_{I \cap s_\ell [I \setminus T]} \mathbf{x}_0\|_1 + \lambda \|\mathbf{a}_0\|_2 \|\mathbf{P}_{I \setminus T} \mathbf{x}_0\|_1 \\ &\quad + \lambda \|\mathbf{C}_{\mathbf{a}_0} \boldsymbol{\iota}\|_2 \sqrt{2p} \|\mathbf{P}_{I \cap s_\ell [I]} \mathbf{x}_0\|_1 \|\boldsymbol{\Delta}\|_{\infty \rightarrow \infty} \\ &\quad + \lambda \|\mathbf{a}_0\|_2 \|\mathbf{x}_0\|_2 \|\boldsymbol{\Delta}\|_2 \sqrt{|J \setminus T|} \\ &\leq C_4 \lambda \left(\sqrt{\tilde{\mu} p^2} \times \tilde{\lambda} n \theta^2 + \lambda \tilde{\lambda} n \theta \right. \\ &\quad \left. + \sqrt{\tilde{\mu} p^2} \times n \theta^2 \times \tilde{\mu} \kappa_I + \sqrt{n\theta} \times \tilde{\mu} \kappa_I \sqrt{\tilde{\lambda} n \theta} \right) \\ &\leq 2C_4 \left(\tilde{\lambda} + \tilde{\mu} \kappa_I \right) \lambda n \theta. \end{aligned} \quad (\text{K.74})$$

Therefore, combining (K.72)-(K.74) we obtain

$$\begin{aligned} & \left\| \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{x}-\mathbf{x}_0} \iota \mathbf{a}_0 - \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1} C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 - \mathbf{a}} \mathbf{x}_0 \right\|_2 \\ & \leq C_5 \left(\tilde{\lambda} + \tilde{\mu} \kappa_I \right) \lambda n \theta. \end{aligned} \quad (\text{K.75})$$

3. (Extract the set J) Lastly, we will further simplify the term with $\mathbf{a} - \mathbf{a}_0$ in (K.75) by extracting the set J :

$$\begin{aligned} & \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1} C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 - \mathbf{a}} \mathbf{x}_0 \\ & = \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (\mathbf{I} + \Delta) C_{\mathbf{a}_0 + (\mathbf{a} - \mathbf{a}_0) J}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \\ & = \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \\ & \quad + \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} \Delta C_{\mathbf{a}_0 J}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \\ & \quad + \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1} C_{\mathbf{a}_0 - \mathbf{a}_0 J}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \\ & \quad - \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_{I \setminus J} C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}), \end{aligned} \quad (\text{K.76})$$

where, the latter terms in (K.76) are bounded as

$$\begin{aligned} & \left\| \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} \Delta C_{\mathbf{a}_0 J}^* C_{\mathbf{x}_0} \iota \right\|_2 \leq \|C_{\mathbf{x}_0} \iota\|_2^2 \|C_{\mathbf{a}_0 J}\|_2^2 \|\Delta\|_2 \leq C_6 \tilde{\mu} \kappa_I n \theta \\ & \left\| \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1} C_{\mathbf{a}_0 - \mathbf{a}_0 J}^* C_{\mathbf{x}_0} \iota \right\|_2 \\ & \leq \|C_{\mathbf{x}_0} \iota\|_2^2 \|C_{\mathbf{a}_0 J}\|_2 \|(C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1}\|_2 \|C_{\mathbf{a}_0 - \mathbf{a}_0 J}\|_2 \leq C_7 \tilde{\mu} \sqrt{\kappa_I} n \theta \\ & \left\| P_{I \setminus J} C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota \right\|_2^2 \leq |I \setminus J| \|\check{\mathbf{a}}_0 * \mathbf{x}_0\|_{\square}^2 \\ & \leq C_8 \tilde{\lambda} n \theta \times \kappa_I \leq C_8 \left(\lambda \kappa_I + \frac{\kappa_I \log n}{\sqrt{n \theta^2}} \right) n \theta, \end{aligned} \quad (\text{K.77})$$

whence we conclude, that since $c_\mu \kappa_I^2 \leq c_\mu$ and $\lambda \kappa_I \leq 5c_\mu$, as long as $c_\mu < \frac{1}{100} \left(\frac{1}{C_6} + \frac{1}{C_7} + \frac{1}{5C_8} \right)$ and $n > 10^6 C_8^2 \theta^{-2} \kappa_I^2 \log^2 n$, we gain:

$$\begin{aligned} & \left\| \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0 J} (C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 J})^{-1} C_{\mathbf{a}_0 J}^* C_{\mathbf{a}_0 - \mathbf{a}} \mathbf{x}_0 \right. \\ & \quad \left. - \iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota (\mathbf{a}_0 - \mathbf{a}) \right\|_2 \\ & \leq \left(\frac{3}{100} + \frac{1}{1000} \right) n \theta \|\mathbf{a}_0 - \mathbf{a}\|_2 \\ & \leq \frac{1}{32} n \theta \|\mathbf{a}_0 - \mathbf{a}\|_2. \end{aligned} \quad (\text{K.78})$$

The claimed result therefore is followed by combining (K.70), (K.75) and (K.78). \blacksquare

The next thing is to show the operator

$$(n\theta)^{-1} \left(\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota \right) \quad (\text{K.79})$$

contracts \mathbf{a} toward \mathbf{a}_0 . We first will show that

$$(n\theta)^{-1} \left(\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota \right) \approx \mathbf{a}_0 \mathbf{a}_0^* \quad (\text{K.80})$$

by seeing $\iota^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} \iota \approx (n\theta) \mathbf{e}_0 \mathbf{e}_0^*$ via sparsity of \mathbf{x}_0 . Finally since the local perturbation on sphere is close to a quadratic function in ℓ^2 -norm of difference, we have

$$|\langle \mathbf{a}_0, \mathbf{a} - \mathbf{a}_0 \rangle| \leq \frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2^2. \quad (\text{K.81})$$

Again, we introduce the following lemma to solidify our claim:

Lemma K.3 (Contraction of \mathbf{a} to \mathbf{a}_0). *Given $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ to be $\tilde{\mu}$ -shift coherent and $\mathbf{x}_0 \sim \text{BG}(\theta) \in \mathbb{R}^n$. There exists some constants C, C', c, c', c_μ such that if $\lambda < c' \tilde{\mu} \kappa_I, \tilde{\mu} \kappa_I^2 \leq c_\mu$ and $n > C\theta^{-2} p^2 \log p$, then with probability at least $1 - c/n$, for every $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$,*

$$\|\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota(\mathbf{a}_0 - \mathbf{a})\|_2 \leq \frac{1}{32} \|\mathbf{a} - \mathbf{a}_0\|_2 n\theta. \quad (\text{K.82})$$

Proof. Since $\mathbb{E} \langle P_I s_i[\mathbf{x}_0], s_j[\mathbf{x}_0] \rangle = 0$ for all $i \neq j$ and set I , we calculate

$$\begin{aligned} \mathbb{E} \left[\iota_{[\pm p]}^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} \iota_{[\pm p]} \right] &= \sum_{i \in [\pm p]} \mathbb{E} [e_i^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} e_i] e_i e_i^* \\ &= \mathbb{E} \|\mathbf{x}_0\|_2^2 e_0 e_0^* + \sum_{i \in [\pm p] \setminus 0} \mathbb{E} \|P_I s_i[\mathbf{x}_0]\|_2^2 e_i e_i^* \\ &= n\theta e_0 e_0^* + n\theta^2 P_{[\pm p] \setminus 0} \\ &= n\theta^2 \mathbf{I} + n\theta(1 - \theta) e_0 e_0^*. \end{aligned} \quad (\text{K.83})$$

whence

$$\begin{aligned} \mathbb{E} [\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota] &= \iota^* C_{\mathbf{a}_0}^* \mathbb{E} [C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0}] C_{\mathbf{a}_0} \iota \\ &= n\theta^2 \iota^* C_{\mathbf{a}_0}^* C_{\mathbf{a}_0} \iota + n\theta(1 - \theta) \mathbf{a}_0 \mathbf{a}_0^*, \end{aligned} \quad (\text{K.84})$$

implying the expectation is a contraction mapping for $\mathbf{a}_0 - \mathbf{a}$ when $c_\mu < \frac{1}{200}$:

$$\begin{aligned} &\|\mathbb{E} [\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota] (\mathbf{a}_0 - \mathbf{a})\|_2 \\ &\leq n\theta^2 \|\iota^* C_{\mathbf{a}_0}^* C_{\mathbf{a}_0} \iota\|_2 \|\mathbf{a}_0 - \mathbf{a}\|_2 + n\theta \|\mathbf{a}_0\|_2 |\langle \mathbf{a}_0, \mathbf{a}_0 - \mathbf{a} \rangle| \\ &\leq n\theta^2 \times 2\tilde{\mu} p \times \|\mathbf{a}_0 - \mathbf{a}\|_2 + \frac{1}{2} n\theta \|\mathbf{a}_0 - \mathbf{a}\|_2^2 \\ &\leq (2c_\mu + \frac{1}{2} c_\mu) \|\mathbf{a}_0 - \mathbf{a}\|_2 n\theta \\ &\leq \frac{1}{64} \|\mathbf{a}_0 - \mathbf{a}\|_2 n\theta. \end{aligned} \quad (\text{K.85})$$

For each entry of $C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0}$, again from [Appendix D](#) we know with probability at least $1 - c/n$:

$$|e_i^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} e_j - \mathbb{E} [e_i^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} e_j]| \leq \begin{cases} C' \sqrt{n\theta \log n} & i = j = 0 \\ C' \sqrt{n\theta^2 \log n} & \text{otherwise} \end{cases}.$$

Thus via Gershgorin disc theorem, when $n > 10^3 C'^2 \theta^{-2} p^2 \log n$:

$$\begin{aligned} \lambda_{\max} \left(\iota_{[\pm p]}^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} \iota_{[\pm p]} - \mathbb{E} [\iota_{[\pm p]}^* C_{\mathbf{x}_0}^* P_I C_{\mathbf{x}_0} \iota_{[\pm p]}] \right) &\leq C' p \sqrt{n\theta^2 \log n} \\ &\leq \frac{1}{64} n\theta^2. \end{aligned} \quad (\text{K.86})$$

Finally we combine [\(K.85\)](#), [\(K.86\)](#) and get

$$\begin{aligned} \|\iota^* C_{\mathbf{x}_0}^* C_{\mathbf{a}_0} P_I C_{\mathbf{a}_0}^* C_{\mathbf{x}_0} \iota(\mathbf{a}_0 - \mathbf{a})\|_2 &\leq \left(\frac{1}{64} n\theta + \frac{1}{64} n\theta^2 \|C_{\mathbf{a}_0} \iota_{[\pm p]}\|_2^2 \right) \|\mathbf{a}_0 - \mathbf{a}\|_2 \\ &\leq \frac{1}{32} \|\mathbf{a}_0 - \mathbf{a}\|_2 n\theta. \end{aligned} \quad (\text{K.87})$$

■

[Lemma K.1-K.3](#) together implies the single iterate contract of alternating minimization contracts \mathbf{a} toward \mathbf{a}_0 . We show it with the following lemma:

Lemma K.4 (Contraction of least square estimate). *Given $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ to be $\tilde{\mu}$ -shift coherent and $\mathbf{x}_0 \sim \text{BG}(\theta) \in \mathbb{R}^n$. There exists some constants C, C', c, c_μ such that if $\tilde{\mu}\kappa_I^2 \leq c_\mu$ and $n > C\theta^{-2}p^2 \log n$, then with probability at least $1 - c/n$, for every λ and \mathbf{a} satisfying*

$$5\tilde{\mu}\kappa_I \geq \lambda \geq 5\kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2, \quad (\text{K.88})$$

and suppose \mathbf{x}^+ has the form of (K.16), then the solution \mathbf{a}^+ to

$$\min_{\mathbf{a}' \in \mathbb{R}^p} \left\{ \|\mathbf{a}' * \mathbf{x}^+ - \mathbf{y}\|_2^2 \right\} \quad (\text{K.89})$$

is unique and satisfies

$$\|\mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{a}^+] - \mathbf{a}_0\|_2 \leq \frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2. \quad (\text{K.90})$$

Proof. Write \mathbf{x} as \mathbf{x}^+ , then

$$\begin{aligned} \lambda_p(\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_x \boldsymbol{\iota}) &= \sigma_{\min}^2(\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota} + \mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota}) \\ &\geq \left[\sigma_{\min}(\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}) - \|\mathbf{C}_{\mathbf{x}-\mathbf{x}_0} \boldsymbol{\iota}\| \right]_+^2 \\ &\geq \left[\sigma_{\min}(\mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}) - 2\sqrt{\kappa_I} \|\mathbf{x} - \mathbf{x}_0\|_2 \right]_+^2 \\ &\geq \left[\frac{2}{3}\sqrt{\theta n} - 8\lambda\sqrt{\kappa_I}\sqrt{\theta n} \right]_+^2 \\ &\geq \frac{1}{2}\theta n, \end{aligned} \quad (\text{K.91})$$

where the fourth inequality is derived from using the upper bound of sparse convolution matrix from Remark D.6, and the last line holds by knowing $\lambda < 5c_\mu\kappa_I^{-1}$. From (K.91) we know the least square problem of (K.89) has unique solution \mathbf{a}^+ , written as

$$\mathbf{a}^+ = (\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_x \boldsymbol{\iota})^{-1} \boldsymbol{\iota} \mathbf{C}_x^* \mathbf{y}, \quad (\text{K.92})$$

whence

$$\begin{aligned} \mathbf{a}^+ - \mathbf{a}_0 &= (\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_x \boldsymbol{\iota})^{-1} (\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_{\mathbf{x}_0} \boldsymbol{\iota}) \mathbf{a}_0 - \mathbf{a}_0 \\ &= (\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_x \boldsymbol{\iota})^{-1} (\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_{\mathbf{x}_0-\mathbf{x}} \boldsymbol{\iota}) \mathbf{a}_0. \end{aligned} \quad (\text{K.93})$$

Combine Lemma K.2 and Lemma K.3, we know

$$\|\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_{\mathbf{x}_0-\mathbf{x}} \boldsymbol{\iota}\|_2 \leq \left(C_1 \lambda (\tilde{\lambda} + \tilde{\mu}\kappa_I) + \frac{1}{16} \|\mathbf{a} - \mathbf{a}_0\|_2 \right) n\theta \quad (\text{K.94})$$

for some constant C_1 . Combine (K.91), (K.93), (K.94) and since $\lambda < \tilde{\mu}\kappa_I$, by letting $c_\mu < \frac{1}{4C_1}$, we gain

$$\begin{aligned} \|\mathbf{a}^+ - \mathbf{a}_0\|_2 &\leq \frac{\|\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_{\mathbf{x}_0-\mathbf{x}} \boldsymbol{\iota}\|_2}{\lambda_p(\boldsymbol{\iota}^* \mathbf{C}_x^* \mathbf{C}_x \boldsymbol{\iota})} \\ &\leq 2C_1 \lambda (\tilde{\lambda} + \tilde{\mu}\kappa_I) + \frac{1}{8} \|\mathbf{a} - \mathbf{a}_0\|_2 \leq \frac{1}{4}. \end{aligned} \quad (\text{K.95})$$

For the final bound,

$$\begin{aligned} \left\| \frac{\mathbf{a}^+}{\|\mathbf{a}^+\|_2} - \mathbf{a}_0 \right\|_2 &\leq \frac{\|\mathbf{a}^+ - \mathbf{a}_0\|_2 + \|\mathbf{a}^+\|_2 - 1}{\|\mathbf{a}^+\|_2} \\ &\leq \frac{2\|\mathbf{a}^+ - \mathbf{a}_0\|_2}{1 - \|\mathbf{a}^+ - \mathbf{a}_0\|_2} \leq \frac{8}{3} \|\mathbf{a}^+ - \mathbf{a}_0\|_2, \end{aligned}$$

$$\leq C_2 \lambda \left(\tilde{\lambda} + \tilde{\mu} \kappa_I \right) + \frac{1}{3} \|\mathbf{a} - \mathbf{a}_0\|_2, \quad (\text{K.96})$$

and since $\lambda > \kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2$, finally we gain

$$\begin{aligned} (\text{K.96}) &\leq C_2 \left(\lambda \kappa_I + \frac{p \kappa_I \log n}{n \theta} + \tilde{\mu} \kappa_I^2 \right) \|\mathbf{a} - \mathbf{a}_0\|_2 + \frac{1}{3} \|\mathbf{a} - \mathbf{a}_0\|_2 \\ &\leq \frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2 \end{aligned} \quad (\text{K.97})$$

as long as $n > 20C_2\theta^{-1}p\kappa_I \log n$ and $c_\mu < \frac{1}{20C_2}$. ■

K.3 Linear convergence of alternating minimization (Proof of Theorem C.2)

In the first two sections we have shown the iterate contract \mathbf{a} toward \mathbf{a}_0 , under our signal assumption. We tie up these result by showing the following theorem which proves that the iterates produced by alternating minimization converge linearly to \mathbf{a}_0 :

Proof. We will prove our claim by induction on k . Clearly, when $k = 0$, we have $5\kappa_I \|\mathbf{a}^{(0)} - \mathbf{a}_0\|_2 \leq \lambda^{(0)} = 5\tilde{\mu}\kappa_I$ and $I^{(0)} = \{i : |s_i[\mathbf{a}^{(0)}]^* \boldsymbol{\iota}^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0| > \lambda^{(0)}\}$. Then for all $|\mathbf{x}_j| > 6\lambda^{(0)}$, we have

$$\begin{aligned} |s_j[\mathbf{a}^{(0)}]^* \mathbf{C}_{\mathbf{a}_0} \mathbf{x}_0| &\geq \left(1 - |\langle \mathbf{a}^{(0)}, \mathbf{a}_0 \rangle|\right) |\mathbf{x}_j| - \left\| \mathbf{P}_{[\pm p] \setminus \{j\}} \mathbf{C}_{\mathbf{a}_0}^* \boldsymbol{\iota} s_j[\mathbf{a}^{(0)}] \right\|_2 \times \sqrt{2} \|\mathbf{x}_0\|_\square \\ &\geq (1 - 2\tilde{\mu}) 6\lambda^{(0)} - 2\tilde{\mu}\sqrt{\kappa_I} \times \sqrt{2\kappa_I} \\ &\geq 5\lambda^{(0)} - 4\lambda^{(0)} \\ &= \lambda^{(0)}. \end{aligned} \quad (\text{K.98})$$

hence $I_{>6\lambda^{(0)}} \subseteq I^{(0)}$, therefore the condition of Lemma K.4 is satisfied, implies (C.32) holds for $k = 0$.

Suppose it is true for $1, 2, \dots, k-1$, such that

$$\kappa_I \|\mathbf{a}^{(k)} - \mathbf{a}_0\|_2 \leq \frac{1}{2} \lambda^{(k-1)} = \lambda^{(k)}, \quad \text{and} \quad I_{>3\lambda^{(k-1)}} \subseteq I^{(k)} \quad (\text{K.99})$$

and since $I_{>6\lambda^{(k)}} = I_{>3\lambda^{(k-1)}} \subseteq I^{(k)}$, we can again apply Lemma K.4, resulting

$$\kappa_I \|\mathbf{a}^{(k+1)} - \mathbf{a}\|_2 \leq \frac{1}{2} \kappa_I \|\mathbf{a}^{(k)} - \mathbf{a}_0\|_2 \leq \frac{1}{2} \lambda^{(k)} \quad (\text{K.100})$$

as claimed. ■

K.4 Supporting lemmas for refinement

The following lemma controls the shift coherence of \mathbf{a} :

Lemma K.5 (Coherence of \mathbf{a} near \mathbf{a}_0). *Suppose that \mathbf{a}_0 is $\tilde{\mu}$ -shift coherent, and $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$. Then*

$$\|\text{off}[\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0}]\|_\infty \leq 2\tilde{\mu} \quad (\text{K.101})$$

$$\|\text{off}[\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}}]\|_\infty \leq 3\tilde{\mu} \quad (\text{K.102})$$

Proof. Notice that for any $\ell \neq 0$, $|\langle \mathbf{a}, s_\ell[\mathbf{a}_0] \rangle| \leq |\langle \mathbf{a}_0, s_\ell[\mathbf{a}_0] \rangle| + |\langle \mathbf{a} - \mathbf{a}_0, s_\ell[\mathbf{a}_0] \rangle| \leq \tilde{\mu} + \|\mathbf{a}_0 - \mathbf{a}\|_2 \leq 2\tilde{\mu}$. Similarly, $|\langle \mathbf{a}, s_\ell[\mathbf{a}] \rangle| \leq |\langle \mathbf{a} - \mathbf{a}_0, s_\ell[\mathbf{a}_0] \rangle| + |\langle \mathbf{a}, s_\ell[\mathbf{a}_0] \rangle| \leq \|\mathbf{a} - \mathbf{a}_0\|_2 + 2\tilde{\mu} \leq 3\tilde{\mu}$, as claimed. ■

From this we obtain the following spectral control on $\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}}$, to simply the notations, we will write

$$\mathbf{C}_{\mathbf{a}I}^* \mathbf{C}_{\mathbf{a}I} = \boldsymbol{\iota}_I^* \mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}} \boldsymbol{\iota}_I = [\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}}]_{I,I} \quad (\text{K.103})$$

in the latter part of this section.

Lemma K.6 (Off-diagonals of $[C_a^* C_a]_{I,I}$). *Suppose that \mathbf{a}_0 is $\tilde{\mu}$ -shift coherent and $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$. Then*

$$\left\| [C_a^* C_a - \mathbf{I}]_{I,I} \right\|_2 \leq 9\kappa_I \tilde{\mu}. \quad (\text{K.104})$$

We prove this lemma by noting that $C_a^* C_a = C_{r_{\mathbf{a},\mathbf{a}}}$ is the convolution matrix associated with the autocorrelation $r_{\mathbf{a},\mathbf{a}}$ of \mathbf{a} . Since $\text{supp}(r_{\mathbf{a},\mathbf{a}}) \subseteq \{-p+1, \dots, p-1\}$ is confined to a (cyclic) stripe of width $2p-1$, we can tightly control the norm of this matrix by dividing it into three block-diagonal submatrices with blocks of size $p \times p$. Formally:

Proof. Divide I into $r = \lceil n/p \rceil$ subsets I_0, \dots, I_{r-1} such that for all $\ell = 0, \dots, r-1$:

$$I_\ell = I \cap \{p\ell, p\ell+1, \dots, p\ell+(p-1)\} = I \cap ([p] + p\ell).$$

Notice that for each ℓ :

$$\text{supp}([C_a^* C_a]_{I_\ell, I}) \subseteq I_\ell \times (I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}),$$

where $\ell+1$ and $\ell-1$ are interpreted cyclically modulo r .

For an arbitrary $\mathbf{v} \in \mathbb{R}^{|I|}$, we calculate

$$\left\| [C_a^* C_a - \mathbf{I}]_{I,I} \mathbf{v} \right\|_2^2 = \sum_{\ell=0}^{r-1} \left\| [C_a^* C_a - \mathbf{I}]_{I_\ell, I} \mathbf{v} \right\|_2^2 \quad (\text{K.105})$$

$$= \sum_{\ell=0}^{r-1} \left\| [C_a^* C_a - \mathbf{I}]_{I_\ell, I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}} \mathbf{v}_{I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}} \right\|_2^2 \quad (\text{K.106})$$

$$\leq \sum_{\ell=0}^{r-1} \left\| [C_a^* C_a - \mathbf{I}]_{I_\ell, I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}} \right\|_F^2 \left\| \mathbf{v}_{I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}} \right\|_2^2 \quad (\text{K.107})$$

$$\leq 3\kappa_I^2 \times (3\tilde{\mu})^2 \times \sum_{\ell=0}^{r-1} \left\| \mathbf{v}_{I_{\ell-1} \uplus I_\ell \uplus I_{\ell+1}} \right\|_2^2 \quad (\text{K.108})$$

$$\leq 3\kappa_I^2 \times 9\tilde{\mu}^2 \times 3 \left\| \mathbf{v} \right\|_2^2, \quad (\text{K.109})$$

giving the claimed result. \blacksquare

As a consequence, we have that

Corollary K.7 (Inverse of $[C_a^* C_a]_{J,J}$). *Suppose that \mathbf{a}_0 is μ -shift coherent, that $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$ and that $\kappa_I \tilde{\mu} < \frac{1}{18}$. Then for every $J \subseteq I$ and any norm $\|\cdot\|_\diamond \in \{\|\cdot\|_{\square \rightarrow \square}, \|\cdot\|_{\infty \rightarrow \infty}, \|\cdot\|_2\}$, we have*

$$\left\| [C_a^* C_a - \mathbf{I}]_{J,J} \right\|_\diamond \leq 9\kappa_I \tilde{\mu} \quad (\text{K.110})$$

$$\left\| [C_a^* C_a]_{J,J}^{-1} - \mathbf{I} \right\|_\diamond \leq 18\kappa_I \tilde{\mu} \quad (\text{K.111})$$

$$\left\| [C_a^* C_a]_{J,J}^{-1} \right\|_\diamond \leq 2. \quad (\text{K.112})$$

Proof. First we prove

$$\left\| [C_a^* C_a - \mathbf{I}]_{J,J} \right\|_2 \leq 9\kappa_I \tilde{\mu}, \quad (\text{K.113})$$

$$\left\| [C_a^* C_a - \mathbf{I}]_{J,J} \right\|_{\infty \rightarrow \infty} \leq 6\kappa_I \tilde{\mu}, \quad (\text{K.114})$$

$$\left\| [C_a^* C_a - \mathbf{I}]_{J,J} \right\|_{\square \rightarrow \square} \leq 6\kappa_I \tilde{\mu} \quad (\text{K.115})$$

Where the first claim follows from [Lemma K.6](#). The second follows by noting that the ℓ^∞ operator norm is the maximum row ℓ^1 norm, and that each row has at most $2\kappa_I$ entries, of size at most $3\tilde{\mu}$. The last follows by noting that

$$\begin{aligned} \left\| [\mathbf{C}_a^* \mathbf{C}_a - \mathbf{I}]_{J,J} \right\|_{\square \rightarrow \square} &\leq \max_{\ell, \ell'} \left\| [\mathbf{C}_a^* \mathbf{C}_a - \mathbf{I}]_{J \cap ([p] + \ell), J \cap ([2p] + \ell')} \right\|_F \\ &\leq 6\kappa_I \tilde{\mu}. \end{aligned} \quad (\text{K.116})$$

Then we prove

$$\begin{aligned} \left\| [\mathbf{C}_a^* \mathbf{C}_a]_{J,J}^{-1} - \mathbf{I} \right\|_2 &\leq 18\kappa_I \tilde{\mu}, \\ \left\| [\mathbf{C}_a^* \mathbf{C}_a]_{J,J}^{-1} - \mathbf{I} \right\|_{\infty \rightarrow \infty} &\leq 12\kappa_I \tilde{\mu}, \\ \left\| [\mathbf{C}_a^* \mathbf{C}_a]_{J,J}^{-1} - \mathbf{I} \right\|_{\square \rightarrow \square} &\leq 12\kappa_I \tilde{\mu}, \end{aligned} \quad (\text{K.117})$$

which are followed from the fact that if $\|\cdot\|_\diamond$ is a matrix norm and $\|\Delta\|_\diamond < 1$, then

$$\left\| (\mathbf{I} + \Delta)^{-1} - \mathbf{I} \right\|_\diamond \leq \frac{\|\Delta\|_\diamond}{1 - \|\Delta\|_\diamond}.$$

Finally, [\(K.112\)](#) follows from the triangle inequality. \blacksquare

Also, we need to bound the convolution of $\mathbf{a}_0 - \mathbf{a}$ with $\|\mathbf{a}_0 - \mathbf{a}\|_2$ requiring for bounds of the lasso solution:

Lemma K.8 (Convolution of $\mathbf{a}_0 - \mathbf{a}$). *Suppose that \mathbf{a}_0 is μ -shift coherent and $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$, then for every $J \subseteq I$,*

$$\left\| [\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0 - \mathbf{a}}]_{J,J} \right\|_{\square \rightarrow \infty} \leq \sqrt{2\kappa_I} \|\mathbf{a} - \mathbf{a}_0\|_2 \quad (\text{K.118})$$

$$\left\| [\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0 - \mathbf{a}}]_{J,J} \right\|_{\square \rightarrow \square} \leq \sqrt{2\kappa_I} \|\mathbf{a} - \mathbf{a}_0\|_2 \quad (\text{K.119})$$

Proof. For the first inequality, we have

$$\begin{aligned} \left\| [\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0 - \mathbf{a}}]_{J,J} \mathbf{v} \right\|_{\square \rightarrow \infty} &= \max_{j \in J, \|\mathbf{v}\|_\square = 1} |\langle s_j[\mathbf{a}], (\mathbf{a}_0 - \mathbf{a}) * \mathbf{v} \rangle| \\ &\leq \max_{j \in [n], \|\mathbf{v}\|_\square = 1} \left\| \mathbf{P}_{[p]+j} [(\mathbf{a}_0 - \mathbf{a}) * \mathbf{v}] \right\|_2 \\ &\leq \|\mathbf{a} - \mathbf{a}_0\|_2 \times \max_{j \in [n], \|\mathbf{v}\|_\square = 1} \left\| \mathbf{P}_{[\pm p]+j} \mathbf{v} \right\|_1 \\ &\leq \sqrt{2\kappa_I} \|\mathbf{a}_0 - \mathbf{a}\|_2 \end{aligned} \quad (\text{K.120})$$

The second inequality is derived by

$$\begin{aligned} \left\| [\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0 - \mathbf{a}}]_{J,J} \right\|_{\square \rightarrow \square} &\leq \max_{\ell, \ell'} \left\| [\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0 - \mathbf{a}}]_{J \cap ([p] + \ell), J \cap ([2p] + \ell')} \right\|_F \\ &\leq \sqrt{2\kappa_I^2 \max_{i,j} |\langle s_i[\mathbf{a}], s_j[\mathbf{a}_0 - \mathbf{a}] \rangle|^2} \\ &\leq \sqrt{2}\kappa_I \|\mathbf{a} - \mathbf{a}_0\|_2, \end{aligned} \quad (\text{K.121})$$

finishing the proof. \blacksquare

Again, using a variant of the argument for [Lemma K.6](#), we have the following:

Lemma K.9 (Off-diagonal of submatrix of $\mathbf{C}_a^* \mathbf{C}_{\mathbf{a}_0}$). *Suppose that \mathbf{a}_0 is μ -shift coherent and $\|\mathbf{a} - \mathbf{a}_0\|_2 \leq \tilde{\mu}$. For any $J \subset I$, if*

$$\kappa_J = \max_{\ell} |J \cap \{\ell, \ell + 1, \dots, \ell + p - 1\}| \quad (\text{K.122})$$

$$\kappa_{I \setminus J} = \max_{\ell} |(I \setminus J) \cap \{\ell, \ell + 1, \dots, \ell + p - 1\}| \quad (\text{K.123})$$

Then

$$\left\| [\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0}]_{J, I \setminus J} \right\|_2 \leq 6\sqrt{\kappa_J \kappa_{I \setminus J}} \tilde{\mu}. \quad (\text{K.124})$$

Proof. Take $r = \lceil n/p \rceil$ and for $\ell = 0, \dots, r-1$, write

$$J_{\ell} = J \cap ([p] + p\ell), \quad L_{\ell} = (I \setminus J) \cap ([p] + p\ell),$$

Take $\mathbf{v} \in \mathbb{R}^{|I \setminus J|}$ arbitrary and notice that

$$\begin{aligned} \left\| [\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0}]_{J, I \setminus J} \mathbf{v} \right\|_2^2 &= \sum_{\ell=0}^{r-1} \left\| [\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0}]_{J_{\ell}, I \setminus J} \mathbf{v} \right\|_2^2 \\ &= \sum_{\ell=0}^{r-1} \left\| [\mathbf{C}_{\mathbf{a}}^* \mathbf{C}_{\mathbf{a}_0}]_{J_{\ell}, L_{\ell-1} \cup L_{\ell} \cup L_{\ell+1}} \mathbf{v}_{L_{\ell-1} \cup L_{\ell} \cup L_{\ell+1}} \right\|_2^2 \\ &\leq 4\tilde{\mu}^2 \times \kappa_J \times 3\kappa_{I \setminus J} \times \sum_{\ell=0}^{r-1} \left\| \mathbf{v}_{L_{\ell-1} \cup L_{\ell} \cup L_{\ell+1}} \right\|_2^2 \\ &\leq 4\tilde{\mu}^2 \times \kappa_J \times 3\kappa_{I \setminus J} \times 3\|\mathbf{v}\|_2^2, \end{aligned} \quad (\text{K.125})$$

giving the result. ■

Lemma K.10 (Perturbation of vector over sphere). *If both \mathbf{a}, \mathbf{a}_0 are unit vectors in inner product space, then*

$$|\langle \mathbf{a}, \mathbf{a} - \mathbf{a}_0 \rangle| \leq \frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2^2. \quad (\text{K.126})$$

Proof. Via simple norm inequalities:

$$\frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2^2 = 1 - \langle \mathbf{a}, \mathbf{a}_0 \rangle = 1 - \langle \mathbf{a}, \mathbf{a}_0 - \mathbf{a} + \mathbf{a} \rangle = \langle \mathbf{a}, \mathbf{a} - \mathbf{a}_0 \rangle > 0 \quad (\text{K.127})$$

Lemma K.11 (Convolution of short and sparse). *Suppose $\delta \in \mathbb{R}^p$, and $\mathbf{v} \in \mathbb{R}^n$ where $\text{supp}(\mathbf{v}) = I$ satisfies*

$$\max_{\ell \in [n]} |I \cap ([p] + \ell)| \leq \kappa \quad (\text{K.128})$$

then

$$\|\delta * \mathbf{v}\|_2 \leq \sqrt{2\kappa} \|\delta\|_2 \|\mathbf{v}\|_2 \quad (\text{K.129})$$

Proof. Since every p -contiguous segment of I has at most κ elements, by splitting $I = I_1 \uplus I_2 \uplus \dots \uplus I_{\kappa} \uplus R$ such that each sets I_i are p -separated:

$$\begin{aligned} I_1 &= \{i_1, i_{\kappa+1}, i_{2\kappa+1}, \dots\} \cap \{0, \dots, n-p-1\}, \\ I_2 &= \{i_2, i_{\kappa+2}, i_{2\kappa+2}, \dots\} \cap \{0, \dots, n-p-1\}, \\ &\vdots \\ I_{\kappa} &= \{i_{\kappa}, i_{2\kappa}, i_{3\kappa}, \dots\} \cap \{0, \dots, n-p-1\}, \end{aligned} \quad (\text{K.130})$$

$$R = I \cap \{n-p, \dots, n-1\}. \quad (\text{K.131})$$

Then the p-separating property gives $\|\delta * P_{I_i} \mathbf{v}\|_2 = \|\delta\|_2 \|P_{I_i} \mathbf{v}\|_2$. Hence:

$$\begin{aligned}
\|\delta * P_I \mathbf{v}\|_2 &= \left\| \sum_{i \in \kappa} \delta * P_{I_i} \mathbf{v} + \delta * P_R \mathbf{v} \right\|_2 \leq \sum_{i \in \kappa} \|\delta * P_{I_i} \mathbf{v}\|_2 + \|\delta * P_R \mathbf{v}\|_2 \\
&= \|\delta\|_2 \sum_{i \in \kappa} \|\mathbf{v}_{I_i}\|_2 + \|\delta\|_2 \|P_R \mathbf{v}\|_1 \\
&\leq \sqrt{\kappa} \|\mathbf{v}_{I_1, \cup, \dots, \cup I_\kappa}\|_2 \|\delta\|_2 + \sqrt{\kappa} \|\mathbf{v}_R\|_2 \|\delta\|_2 \\
&\leq \sqrt{2\kappa} \|\mathbf{v}\|_2 \|\delta\|_2,
\end{aligned} \tag{K.132}$$

where the last two inequalities were coming from Cauchy-Schwartz. \blacksquare

L Finite sample approximation

In this section we collect several major components of proof about large sample deviation. In particular, the concentration for shift space gradient $\chi(\beta)_i$, shift space Hessian diagonals $\|P_{I(\mathbf{a})} s_{-i}[\mathbf{x}_0]\|_2$, and the set of gradients discontinuity entries $|J_B(\mathbf{a})|$.

L.1 Proof of Corollary F.4

Proof. 1. (ε -net) Write \mathbf{x} as \mathbf{x}_0 and $\|\beta\|_2 = \eta$ through out this proof, firstly from Definition E.1 for every $\mathbf{a} \in \cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$, we know $\eta \leq 1 + c_\mu + \frac{c_\mu}{\sqrt{\theta k \log \theta^{-1}}} \leq \sqrt{p}$. Define $\varepsilon = \frac{c_2}{2n^{3/2} p^{3/2}}$ and consider the ε -net \mathcal{N}_ε for sphere of radius η . From Lemma N.5 we know for any $c_2 < 1$:

$$|\mathcal{N}_\varepsilon| \leq \left(\frac{3\eta}{\varepsilon}\right)^{2p} \leq \left(\frac{3n^{3/2} p^2}{c_2}\right)^{2p} \leq \left(\frac{3np^2}{c_2}\right)^{3p} \tag{L.1}$$

for each $i \in [n]$ define such net as $\mathcal{N}_{\varepsilon, i}$, and define an event such that all center of subsets in $\mathcal{N}_{\varepsilon, i}$ are being well-behaved:

$$\mathcal{E}_{\text{Net}} := \left\{ \forall i \in [n], \quad \sigma_i n^{-1} \chi[\beta_\varepsilon]_i - \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta_\varepsilon]}_i < \frac{c_1 \theta}{p^{3/2}} \quad \forall \beta_\varepsilon \in \mathcal{N}_{\varepsilon, i} \right\} \tag{L.2}$$

2. (Lipschitz constant) The Lipschitz constant L of $\chi[\cdot]_i$ w.r.t β is bounded in terms of \mathbf{x} regardless of entry i :

$$\begin{aligned}
|\chi[\beta]_i - \chi[\beta']_i| &\leq \left| e_i^* \check{C}_x \mathcal{S}_\lambda [\check{C}_x \beta] - e_i^* \check{C}_x \mathcal{S}_\lambda [\check{C}_x \beta'] \right| \\
&\leq \|\mathbf{x}\|_2 \left\| \mathcal{S}_\lambda [\check{C}_x \beta] - \mathcal{S}_\lambda [\check{C}_x \beta'] \right\|_2 \\
&\leq \|\mathbf{x}\|_2 \sqrt{\sum_{j \in [n]} \left| \mathcal{S}_\lambda [\check{C}_x \beta]_j - \mathcal{S}_\lambda [\check{C}_x \beta']_j \right|^2} \\
&\leq \|\mathbf{x}\|_2 \left\| \check{C}_x \beta - \check{C}_x \beta' \right\|_2 \\
&\leq \|\mathbf{x}\|_2 \cdot \|\mathbf{x}\|_1 \cdot \|\beta - \beta'\|_2 =: L \|\beta - \beta'\|_2
\end{aligned} \tag{L.3}$$

Define the event that $\chi[\beta]_i$ that has small Lipschitz constant as

$$\mathcal{E}_{\text{Lip}} := \left\{ L < 2n^{3/2} \theta \right\} \tag{L.4}$$

on the event \mathcal{E}_{Lip} , for every points in $\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and $i \in [n]$, there exists some $\beta_\varepsilon \in \mathcal{N}_{\varepsilon, i}$ such that

$$\left| \left(\sigma_i n^{-1} \chi[\beta]_i - \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta]_i} \right) - \left(\sigma_i n^{-1} \chi[\beta_\varepsilon]_i - \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta_\varepsilon]_i} \right) \right| \leq 2L\varepsilon \leq \frac{c_2 \theta}{p^{3/2}} \quad (\text{L.5})$$

On event $\mathcal{E}_{\text{Lip}} \cap \mathcal{E}_{\text{Net}}$, (L.2), (L.5) implies $\chi[\beta]$ is well concentrated entrywise and anywhere in $\cup_{|\tau| \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$:

$$\left| \sigma_i n^{-1} \chi[\beta]_i - \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta]_i} \right| \leq \frac{(c_1 + c_2) \theta}{p^{3/2}}, \quad \forall \mathbf{a} \in \cup_{k \leq k} \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu)), \quad \forall i \in [n] \quad (\text{L.6})$$

as desired, where, using Lemma D.2,

$$\mathbb{P} [\mathcal{E}_{\text{Lip}}^c] \leq \mathbb{P} \left[\|\mathbf{x}\|_2^2 > 2n\theta \right] \leq 1/n; \quad (\text{L.7})$$

and using union bound,

$$\begin{aligned} \mathbb{P} [\mathcal{E}_{\text{Net}}^c] &\leq \mathbb{P} \left[\max_{\substack{\mathbf{a}_\varepsilon \in \mathcal{N}_{\varepsilon, i} \\ i \in [n]}} \sigma_i n^{-1} \chi[\beta_\varepsilon]_i - \sigma_i n^{-1} \overline{\mathbb{E} \chi[\beta_\varepsilon]_i} > \frac{c_1 \theta}{p^{3/2}} \right] \\ &\leq n |\mathcal{N}_\varepsilon| \mathbb{P} \left[\sigma_0 n^{-1} \chi[\beta_\varepsilon]_0 - \sigma_0 n^{-1} \overline{\mathbb{E} \chi[\beta_\varepsilon]_0} > \frac{c_1 \theta}{p^{3/2}} \right]. \end{aligned} \quad (\text{L.8})$$

3. (Bound $\mathbb{P} [\mathcal{E}_{\text{Net}}^c]$) Wlog write $n = t \cdot (2p)$ for some integer t and $2p \geq 4p_0 - 3$ and replace \mathbf{x}_0 with \mathbf{x} . Observe that $\mathbf{Z}_j(\beta)$ from (F.9) is independent of $\mathbf{Z}_{j+2p}(\beta)$ for all $j \in [n]$ while all \mathbf{Z}_j are identical distributed. We write $\chi[\beta]_0$ as sum of iid r.v.s. as

$$\chi[\beta]_0 = \sum_{j \in [n]} \mathbf{Z}_j(\beta) = \sum_{k \in [2p]} \left(\sum_{t=0}^{n/2p-1} \mathbf{Z}_{k+2tp}(\beta) \right)$$

wlog let $\sigma_0 = 1$ and split the independent r.v.s, write $\mathbb{E} \mathbf{Z}_0 = \mathbb{E} \mathbf{Z}$, bound the tail probability of $\chi[\beta]_0$ as

$$\mathbb{P} \left[n^{-1} \chi[\beta]_0 > n^{-1} \overline{\mathbb{E} \chi(\beta)_0} + \frac{c_1 \theta}{p^{3/2}} \right] \leq 2p \cdot \mathbb{P} \left[\sum_{t=0}^{n/2p-1} \mathbf{Z}_{2tp}(\beta) > \frac{n}{2p} \mathbb{E} \mathbf{Z}(\beta) + \frac{c_1 n \theta}{2p^{5/2}} \right] \quad (\text{L.9})$$

The moments of \mathbf{Z}_0 can be bounded by using $|\mathbf{Z}_0(\beta)| \leq |\mathbf{x}_0| |\beta_0 \mathbf{x}_0 + \mathbf{s}_0| \leq \beta_0 \mathbf{x}_0^2 + |\mathbf{x}_0| |\mathbf{s}_0|$ where $\mathbf{s}_0 = \sum_{\ell \neq 0} \mathbf{x}_\ell \beta_\ell$, write $\mathbf{x} = \boldsymbol{\omega} \circ \mathbf{g} \sim_{\text{i.i.d.}} \text{BG}(\theta)$. For the 2-norm we know

$$\mathbb{E} |\mathbf{s}_0|^2 = \mathbb{E} \left| \sum_{\ell} \mathbf{x}_\ell \beta_\ell \right|^2 \leq \theta \|\beta\|_2^2 \leq \theta \left(1 + c_\mu + \frac{c_\mu}{\theta k^2} \right) \leq \frac{1}{2} \quad (\text{L.10})$$

As for the q -norm, use the moment generating function bound, such that for all $t \geq 0$:

$$\begin{aligned} \mathbb{E} |\mathbf{s}_0|^q &\leq q! t^{-q} \mathbb{E} \exp [t |\mathbf{s}_0|] \leq q! t^{-q} \prod_{\ell} \mathbb{E}_{\boldsymbol{\omega}_\ell, \mathbf{g}_\ell} \exp [t \boldsymbol{\omega}_\ell |\mathbf{g}_\ell| |\beta_\ell|] \\ &\leq 2q! t^{-q} \prod_{\ell} \mathbb{E}_{\boldsymbol{\omega}_\ell} \exp [\boldsymbol{\omega}_\ell t^2 \beta_\ell^2 / 2] \\ &\leq 2q! t^{-q} \prod_{\ell} (1 - \theta + \theta \exp [t^2 \beta_\ell^2 / 2]) \end{aligned} \quad (\text{L.11})$$

notice that the entrywise twice derivative of (L.11) w.r.t. β_ℓ^2 's are always positive, this function is convex for all β_ℓ^2 . Constrain on the polytope $\sum_\ell \beta_\ell^2 \leq \|\beta\|_2^2$, the maximizer of (L.11) w.r.t. β_ℓ^2 's occurs and a vertex point where $\beta_0^2 = \|\beta\|_2^2$. Thus

$$(L.11) \leq 2q!t^{-q} \left(1 - \theta + \theta \exp \left[t^2 \|\beta\|_2^2 / 2 \right] \right) \prod_{\ell \neq 0} (1 - \theta + \theta e^0) \leq 2q!t^{-q} (1 + \theta \exp[\|\beta\|_2^2 t^2 / 2]).$$

Choose $t = \sqrt{q} / \|\beta\|_2$, use $q!! > (q!/2) \cdot (e/q)^{q/2}$, we have

$$\mathbb{E} |\mathbf{s}_0|^q \leq 2q!q^{-q/2} \|\beta\|_2^q (1 + \theta \exp[q/2]) \leq 8 \|\beta\|_2^q \max \{ e^{-q/2}, \theta \} q!! \quad (L.12)$$

Apply Jensen's inequality $\left(\sum_{i=1}^N z_i \right)^q \leq N^{q-1} \sum_{i=1}^N z_i^q$, use Gaussian moment Lemma N.2, (L.10) and (L.12), obtain for $q \geq 3$,

$$\begin{aligned} \mathbb{E} Z(\beta)^2 &\leq \mathbb{E} (\beta_0 \mathbf{x}_0^2 + |\mathbf{x}_0| |\mathbf{s}_0|)^2 \leq 2\mathbb{E} [\beta_0^2 \mathbf{x}_0^4 + \mathbf{x}_0^2 \mathbf{s}_0^2] \leq 6\theta + 2\theta^2 \|\beta\|_2^2 \leq 7\theta, \\ \mathbb{E} Z(\beta)^q &\leq \mathbb{E} (\beta_0 \mathbf{x}_0^2 + |\mathbf{x}_0| |\mathbf{s}_0|)^q \leq 2^{q-1} \left(\mathbb{E} \mathbf{x}_0^{2q} + \mathbb{E} |\mathbf{x}_0|^q \mathbb{E} |\mathbf{s}_0|^q \right) \\ &\leq \theta 2^{q-1} (2q-1)!! + \theta 2^{q-1} (q-1)!! \left(8 \|\beta\|_2^q \max \{ e^{-q/2}, \theta \} q!! \right) \\ &\leq \theta 4^q q! + \theta 2^q \|\beta\|_2^q q!. \end{aligned}$$

Thus, recall that $\|\beta\|_2 = \eta$, use $(\sigma^2, R) = (8\theta\eta^2, 4\eta)$, from (L.8)-(L.9), apply Bernstein inequality Lemma N.4 with $n \geq Cp^5\theta^{-2} \log p$, and $c_1, c_2 \in [0, 1]$ we have

$$\begin{aligned} \mathbb{P} [\mathcal{E}_{\text{Net}}^c] &\leq 2np |\mathcal{N}_\varepsilon| \cdot \mathbb{P} \left[\sum_{t=0}^{n/2p-1} \mathbf{Z}_{2tp}(\beta) > \frac{n}{2p} \mathbb{E} Z(\beta) + \frac{c_1 n \theta}{2p^{5/2}} \right] \\ &\leq 2np \left(\frac{3np^2}{c_2} \right)^{3p} \exp \left(\frac{-(c_1 n \theta / 2p^{5/2})^2}{16n\theta\eta^2/2p + 8\eta c_1 n \theta / 2p^{5/2}} \right) \\ &\leq \exp \left(4p \log \left(\frac{3np^2}{c_2} \right) - \frac{(c_1 n \theta / 2p^{5/2})^2}{16n\theta\eta^2/p} \right) \\ &\leq \exp \left(4p \log \left(\frac{3np^2}{c_2} \right) - \frac{c_1^2 n \theta^2}{64p^4} \right) \\ &\leq \exp \left(\frac{-c_1^2 n \theta^2}{100p^4} \right) \leq \frac{1}{n} \quad (L.13) \end{aligned}$$

when $\frac{C}{\log C} > \frac{10^5}{c_1^2 c_2}$. The proof of lower bound and negative β_0 is derived in the same manner. \blacksquare

L.2 Proof of Corollary G.3

Proof. Write \mathbf{x} as \mathbf{x}_0 though our this proof. Write $\beta_i \mathbf{x}_j + \mathbf{s}_j = \sum_{\ell \in [\pm p]} \beta_\ell \mathbf{x}_{\ell-i+j} = \langle \beta, \mathbf{x}_{[\pm p]-i+j} \rangle$, and the support w.r.t. some \mathbf{a} as $I(\beta)$. Define the random variable $\mathbf{Z}_{ij}(\beta)$ as

$$\|\mathbf{P}_{I(\beta)} \mathbf{s}_{-i}[\mathbf{x}]\|_2^2 = \sum_{j \in [n]} \mathbf{x}_j^2 \mathbf{1}_{\{|\langle \beta, \mathbf{x}_{[\pm p]-i+j} \rangle| > \lambda\}} =: \sum_{j \in [n]} \mathbf{Z}_{ij}(\beta) \quad (L.14)$$

and define $\{\bar{\mathbf{Z}}_{ij}(\beta)\}_{j \in [n]}$ that are independent r.v.s. and as a upper bounding function of $\mathbf{Z}_{ij}(\beta)$ as

$$\bar{\mathbf{Z}}_{ij}(\beta) := \begin{cases} \mathbf{x}_j^2, & |\langle \beta, \mathbf{x}_{[\pm p]-i+j} \rangle| > \lambda \\ 0, & |\langle \beta, \mathbf{x}_{[\pm p]-i+j} \rangle| < \lambda/2, \\ \frac{\mathbf{x}_j^2}{\lambda/2} (|\langle \beta, \mathbf{x}_{[\pm p]-i+j} \rangle| - \lambda/2), & \text{otherwise} \end{cases} \quad (L.15)$$

Similar to proof of [Corollary F.4](#). Let $\|\beta\|_2 \leq \eta \leq \sqrt{p}$. Define $\varepsilon = \frac{c'_2 \lambda}{24np\sqrt{p\theta \log n \log \theta^{-1}}}$ for some $c'_2 > 0$ and consider the ε -net \mathcal{N}_ε for sphere of radius η . From [Lemma N.5](#) we know

$$|\mathcal{N}_\varepsilon| \leq \left(\frac{3\eta}{\varepsilon}\right)^{2p} \leq \left(\frac{72}{c'_2 c_\lambda} np^2 \sqrt{\theta |\tau| \log n \log \theta^{-1}}\right)^{2p} \leq \left(\frac{72}{c'_2 c_\lambda} np^2 \log n\right)^{2p}, \quad (\text{L.16})$$

for each $i \in [n]$ define such net as $\mathcal{N}_{\varepsilon,i}$, and define an event such that all center of subsets in $\mathcal{N}_{\varepsilon,i}$ are being well-behaved:

$$\mathcal{E}_{\text{Net}} := \left\{ \forall i \in [n], \left| n^{-1} \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta_\varepsilon) - \mathbb{E} \bar{\mathbf{Z}}_i(\beta_\varepsilon) \right| \leq \frac{c'_1 \theta}{p} \quad \forall \beta_\varepsilon \in \mathcal{N}_{\varepsilon,i} \right\}, \quad (\text{L.17})$$

Also, $\sum_j \bar{\mathbf{Z}}_{ij}(\beta)$ is a Lipchitz function over β for every $i \in [n]$ as

$$\begin{aligned} \left| \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta) - \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta') \right| &\leq \sum_{j \in [n]} \frac{\mathbf{x}_j^2}{\lambda/2} |\langle \beta - \beta', \mathbf{x}_{[\pm p] - i + j} \rangle| \\ &\leq \sum_{j \in [n]} \frac{\mathbf{x}_j^2 \|\mathbf{x}_{[\pm p] - i + j}\|_2}{\lambda/2} \|\beta - \beta'\|_2, \\ &\leq \frac{1}{\lambda/2} \|\mathbf{x}\|_2^2 \cdot \max_{j \in [n]} \|\mathbf{x}_{[\pm p] + j}\|_2 \cdot \|\beta - \beta'\|_2 \\ &:= L \|\beta - \beta'\|_2, \end{aligned} \quad (\text{L.18})$$

and define event \mathcal{E}_{Lip} such that the Lipchitz constant is bounded as

$$\mathcal{E}_{\text{Lip}} := \left\{ L \leq 12n\theta \sqrt{p\theta \log n \log \theta^{-1}} \lambda^{-1} \right\}, \quad (\text{L.19})$$

then on event \mathcal{E}_{Lip} , for any points β in $\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and $i \in [n]$, there exists some β_ε in $\mathcal{N}_{\varepsilon,i}$ with $\|\beta - \beta_\varepsilon\|_2 \leq \varepsilon$, and thus

$$\left| \left(n^{-1} \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta) - \mathbb{E} \bar{\mathbf{Z}}_i(\beta) \right) - \left(n^{-1} \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta_\varepsilon) - \mathbb{E} \bar{\mathbf{Z}}_i(\beta_\varepsilon) \right) \right| \leq 2L\varepsilon \leq \frac{c'_2 \theta}{p}. \quad (\text{L.20})$$

On event $\mathcal{E}_{\text{Lip}} \cap \mathcal{E}_{\text{Net}}$, from [\(L.17\)](#), [\(L.20\)](#), we can conclude that for all $\beta \in \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and $i \in [n]$ that:

$$\begin{aligned} n^{-1} \|\mathbf{P}_{I(\beta) s_{-i}}[\mathbf{x}_0]\|_2^2 - n^{-1} \mathbb{E} \|\mathbf{P}_{I(\beta) s_{-i}}[\mathbf{x}_0]\|_2^2 &\leq n^{-1} \sum_{j \in [n]} \bar{\mathbf{Z}}_{ij}(\beta) - \mathbb{E} \bar{\mathbf{Z}}_i(\beta) \\ &\leq \frac{(c'_1 + c'_2)\theta}{p} \end{aligned} \quad (\text{L.21})$$

as desired, where the error probability of $\mathcal{E}_{\text{Lip}}^c$ is bounded using [Lemma D.2](#) and [Lemma D.3](#), which give

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{\text{Lip}}^c] &\leq \mathbb{P}[\|\mathbf{x}\|_2^2 > 2n\theta] + \mathbb{P}\left[\max_{j \in [n]} \|\mathbf{x}_{[\pm p] + j}\|_2 > 3\sqrt{p\theta \log n \log \theta^{-1}}\right] \\ &\leq 3/n, \end{aligned} \quad (\text{L.22})$$

when $n > 10^3 \theta^{-1}$. As for $\mathcal{E}_{\text{Net}}^c$ use union bound and split the r.v.s since $\mathbf{Z}_j, \mathbf{Z}_{j+2p}$ are independent for all j :

$$\mathbb{P}[\mathcal{E}_{\text{Net}}^c] \leq 2np \cdot |\mathcal{N}_\varepsilon| \cdot \mathbb{P}\left[\left| \sum_k \bar{\mathbf{Z}}_{i,2kj}(\beta) - \frac{n}{2p} \mathbb{E} \bar{\mathbf{Z}}_i(\beta) \right| \geq \frac{c'_1 n \theta}{2p^2} \right].$$

Now we calculate the variance and L^q -norm of $\sum_k \bar{\mathbf{Z}}_{i,2kj}$ for $q \geq 3$:

$$\begin{cases} \mathbb{E} \bar{\mathbf{Z}}_{i,j}^2 \leq \mathbb{E} \mathbf{x}_j^4 \leq 3\theta \\ \mathbb{E} \bar{\mathbf{Z}}_{i,j}^q \leq \mathbb{E} \mathbf{x}_j^{2q} \leq \theta(2q-1)!! \leq \frac{1}{2} \cdot (3\theta) \cdot 2^{q-2} q! \end{cases} \quad (\text{L.23})$$

and apply Bernstein inequality with $(\sigma^2, R) = (3\theta, 2)$, then use $n \geq Cp^4\theta^{-1} \log p$ and $c'_1, c'_2 < 1$ to obtain

$$\begin{aligned} & 2np |\mathcal{N}_\varepsilon| \mathbb{P} \left[\left| \sum_k^{n/2p} \bar{\mathbf{Z}}_{i,2kj}(\boldsymbol{\beta}) - \frac{n}{2p^2} \mathbb{E} \bar{\mathbf{Z}}_i \right| \geq \frac{c'_1 n \theta}{2p^2} \right] \\ & \leq \exp \left[\log(2np) + 2p \log \left(\frac{72}{c'_2 c_\lambda} np^2 \log n \right) - \frac{(c'_1 n \theta / 2p^2)^2}{6n\theta/2p + 4c'_1 n \theta / 2p^2} \right] \\ & \leq \exp \left[3p \log \left(\frac{72}{c'_2 c_\lambda} np^2 \log n \right) - \frac{c'_1{}^2 n \theta}{24p^3} \right] \\ & \leq \exp[-c'_1{}^2 n \theta / (50p^3)] \leq 1/n, \end{aligned} \quad (\text{L.24})$$

where the last two inequalities holds when $\frac{C}{\log C} \geq \frac{10^5}{c_1{}^2 c_2 c_\lambda}$. The other side of inequality of (G.9) can be derived by defining $\underline{\mathbf{Z}}_{ij}$ as

$$\underline{\mathbf{Z}}_{ij}(\boldsymbol{\beta}) := \begin{cases} \mathbf{x}_j^2, & |\langle \boldsymbol{\beta}, \mathbf{x}_{[\pm p]-i+j} \rangle| > 3\lambda/2 \\ 0, & |\langle \boldsymbol{\beta}, \mathbf{x}_{[\pm p]-i+j} \rangle| < \lambda \\ \frac{\mathbf{x}_j^2}{\lambda/2} (|\langle \boldsymbol{\beta}, \mathbf{x}_{[\pm p]-i+j} \rangle| - \lambda), & \text{otherwise} \end{cases}, \quad (\text{L.25})$$

and define $\mathcal{E}_{\text{Net}}, \mathcal{E}_{\text{Lip}}$ similarly, such that on intersection of these events,

$$\begin{aligned} n^{-1} \|\mathbf{P}_{I(\boldsymbol{\beta})} s_{-i}[\mathbf{x}]\|_2^2 - n^{-1} \mathbb{E} \|\mathbf{P}_{I(\boldsymbol{\beta})} s_{-i}[\mathbf{x}]\|_2^2 & \geq n^{-1} \sum_{j \in [n]} \underline{\mathbf{Z}}_{ij}(\boldsymbol{\beta}) - \mathbb{E} \underline{\mathbf{Z}}_i(\boldsymbol{\beta}) \\ & \geq \frac{(c'_1 + c'_2)\theta}{p} \end{aligned} \quad (\text{L.26})$$

as desired. ■

L.3 Proof of Lemma H.5

Proof. 1. (Expectation upper bound) We will write \mathbf{x} as \mathbf{x}_0 . Similar to proof of Corollary F.4 let $\|\boldsymbol{\beta}\|_2 \leq \eta \leq \sqrt{p}$. For each $i \in [n]$, define the random variable

$$\mathbf{X}_i(\boldsymbol{\beta}) = \mathbf{1}_{\{|\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle - \lambda| \leq B\}} + \mathbf{1}_{\{|\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle + \lambda| \leq B\}}, \quad (\text{L.27})$$

then number of indices for vector $\mathbf{x} * \check{\boldsymbol{\beta}}$ that are within B of $\pm\lambda$ is a random variable $\sum_{i \in [n]} \mathbf{X}_i(\boldsymbol{\beta})$. For each of the $\mathbf{X}_i(\boldsymbol{\beta})$'s consider an upper bound $\bar{\mathbf{X}}_i(\boldsymbol{\beta})$ defined as

$$\bar{\mathbf{X}}_i(\boldsymbol{\beta}) = \begin{cases} \frac{1}{M} (|\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle| - (\lambda - B - M)) & |\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle| \in [\lambda - B - M, \lambda - B] \\ 1 & |\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle| \in [\lambda - B, \lambda + B] \\ \frac{1}{M} ((\lambda + B + M) - |\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle|) & |\langle s_i[\mathbf{x}], \boldsymbol{\beta} \rangle| \in [\lambda + B, \lambda + B + M] \\ 0 & \text{else} \end{cases} \quad (\text{L.28})$$

where $B < M = c\lambda\theta^2 / (p \log n) \leq \lambda/4$ for some constant $0 < c < 1$.

Notice that $\mathbf{x} \sim_{\text{i.i.d.}} \text{BG}(\theta)$ is equal in distribution to $\mathbf{P}_{I(\mathbf{a})}\mathbf{g}$, where $\mathbf{g} \sim_{\text{i.i.d.}} \mathcal{N}(0, 1)$, and $I(\mathbf{a}) \subseteq [n]$ is an independent Bernoulli subset. Conditioned on $I(\mathbf{a})$, $\langle \mathbf{x}, \boldsymbol{\beta} \rangle = \langle \mathbf{g}, \mathbf{P}_{I(\mathbf{a})}\boldsymbol{\beta} \rangle \sim \mathcal{N}(0, \|\mathbf{P}_{I(\mathbf{a})}\boldsymbol{\beta}\|_2^2)$. For all realizations of $I(\mathbf{a})$, the variance $\|\mathbf{P}_{I(\mathbf{a})}\boldsymbol{\beta}\|_2^2$ is bounded by $\|\mathbf{P}_{I(\mathbf{a})}\boldsymbol{\beta}\|_2^2 \leq \|\boldsymbol{\beta}\|_2^2 \leq p$. Using these observations, and letting $f_\sigma(t) = (\sqrt{2\pi}\sigma)^{-1} \exp(-t^2/2\sigma^2)$ denote the pdf of an $\mathcal{N}(0, \sigma^2)$ random variable, the expectation of $\sum_i \overline{\mathbf{X}}_i(\boldsymbol{\beta})$ can be upper bounded as

$$\begin{aligned} \sum_{i \in [n]} \mathbb{E} [\overline{\mathbf{X}}_i(\boldsymbol{\beta})] &\leq (2n) \cdot \mathbb{P}[\langle \mathbf{x}, \boldsymbol{\beta} \rangle \in [\lambda - B - M, \lambda + B + M]] \\ &\leq (2n) \cdot 2(B + M) \sup_{\sigma^2 \in (0, p]} \max_{t \in [\lambda - B - M, \lambda + B + M]} f_\sigma(t) \\ &\leq 4n(B + M) \sup_{\sigma^2 \in (0, p]} f_\sigma(\lambda - B - M) \\ &\leq 4n(B + M) \sup_{\sigma^2 \in (0, p]} f_\sigma(\lambda/2). \end{aligned} \quad (\text{L.29})$$

Notice that

$$\frac{d}{d\sigma} f_\sigma\left(\frac{\lambda}{2}\right) = \frac{d}{d\sigma} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\lambda^2}{8\sigma^2}\right) = \frac{\lambda^2 - 4\sigma^2}{4\sqrt{2\pi}\sigma^4} \exp\left(-\frac{\lambda^2}{8\sigma^2}\right),$$

and hence $f_\sigma(\lambda/2)$ is maximized at either $\sigma^2 = 0$, $\sigma^2 = p$ or $\sigma^2 = \lambda^2/4$. Comparing values at these points, we obtain that

$$\sup_{\sigma^2 \in (0, p]} f_\sigma(\lambda/2) \leq f_{\lambda/2}(\lambda/2) \leq \frac{1}{\sqrt{2\pi}(\lambda/2)} \exp\left(-\frac{1}{2}\right) \leq \frac{1}{2\lambda}, \quad (\text{L.30})$$

whence, by letting $B \leq c\lambda\theta^2/(p \log n)$, the upper bound of expectation become:

$$\sum_{i \in [n]} \mathbb{E} [\overline{\mathbf{X}}_i(\boldsymbol{\beta})] \leq \frac{4n}{2\lambda} (B + M) \leq \frac{4cn\theta^2}{p \log n} =: n\mathbb{E}\overline{\mathbf{X}}(\boldsymbol{\beta}). \quad (\text{L.31})$$

2. (ε -net) Define $\varepsilon = \frac{c^2\lambda\theta^{3.5}}{3p^{2.5}\log^{2.5}n \log^{0.5}\theta^{-1}}$. Write $\lambda = c_\lambda/\sqrt{|\tau|}$ and consider the ε -net \mathcal{N}_ε for sphere of radius $\eta \leq \sqrt{p}$. From [Lemma N.5](#) we know

$$|\mathcal{N}_\varepsilon| \leq \left(\frac{3\eta}{\varepsilon}\right)^{2p} \leq \left(\frac{81|\tau|p^6 \log^5 n \log \theta^{-1}}{c^4 c_\lambda^2 \theta^7}\right)^p \leq \left(\frac{2p \log n}{c \cdot c_\lambda}\right)^{13p} \quad (\text{L.32})$$

and define an event such that all center of subsets in \mathcal{N}_ε are being well-behaved:

$$\mathcal{E}_{\text{Net}} := \left\{ \sum_{i \in [n]} \overline{\mathbf{X}}_i(\boldsymbol{\beta}_\varepsilon) - n\mathbb{E}\overline{\mathbf{X}}(\boldsymbol{\beta}_\varepsilon) < \frac{18cn\theta^2}{p \log n} \quad \forall \boldsymbol{\beta}_\varepsilon \in \mathcal{N}_\varepsilon, \right\} \quad (\text{L.33})$$

3. (Lipschitz constant) Furthermore, the function $\sum_i \overline{\mathbf{X}}_i(\boldsymbol{\beta})$ is Lipchitz over $\boldsymbol{\beta}$ such that

$$\begin{aligned} \left| \sum_{i \in [n]} \overline{\mathbf{X}}_i(\boldsymbol{\beta}) - \sum_{i \in [n]} \overline{\mathbf{X}}_i(\boldsymbol{\beta}') \right| &\leq \sum_{i \in [n]} \frac{1}{M} |\langle s_i[\mathbf{x}], \boldsymbol{\beta} - \boldsymbol{\beta}' \rangle| \\ &\leq \frac{n}{M} \max_{i \in [n]} \|\mathbf{P}_{[\pm p]+i}\mathbf{x}\|_2 \|\boldsymbol{\beta} - \boldsymbol{\beta}'\|_2 \end{aligned}$$

$$=: L \|\beta - \beta'\|_2$$

define the set \mathcal{N}_ε where Lipschitz constant is well bounded:

$$\mathcal{E}_{\text{Lip}} := \left\{ L \leq \frac{3n\sqrt{p\theta \log n \log \theta^{-1}}}{M} \right\},$$

then on event \mathcal{E}_{Lip} , for every β in $\mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$, there exists some β_ε in $\mathcal{N}_{\varepsilon,i}$ with $\|\beta - \beta_\varepsilon\|_2 \leq \varepsilon$, thus

$$\left| \left(\sum_{i \in [n]} \bar{\mathbf{X}}_i(\beta) - n\mathbb{E}\bar{\mathbf{X}}(\beta) \right) - \left(\sum_{i \in [n]} \bar{\mathbf{X}}_i(\beta_\varepsilon) - n\mathbb{E}\bar{\mathbf{X}}(\beta_\varepsilon) \right) \right| \leq 2L\varepsilon \leq \frac{2cn\theta^2}{p \log n}. \quad (\text{L.34})$$

On event $\mathcal{E}_{\text{Lip}} \cap \mathcal{E}_{\text{Net}}$, from (L.31), (L.33) and (L.34), we can conclude that for every $\beta \in \mathfrak{R}(\mathcal{S}_\tau, \gamma(c_\mu))$ and $i \in [n]$,

$$\sum_{i \in [n]} \bar{\mathbf{X}}_i(\beta) \leq \frac{24cn\theta^2}{p \log n} \quad (\text{L.35})$$

as desired, where the error probability of $\mathcal{E}_{\text{Lip}}^c$ is bounded using Lemma D.3, which gives

$$\mathbb{P}[\mathcal{E}_{\text{Lip}}^c] \leq \mathbb{P} \left[\max_{j \in [n]} \|\mathbf{x}_{[\pm p]+j}\|_2 > 3\sqrt{p\theta \log n \log \theta^{-1}} \right] \leq 2/n, \quad (\text{L.36})$$

4. (Bound $\mathbb{P}[\mathcal{E}_{\text{Net}}^c]$) Wlog let us assume that $2p$ divides n . By applying union bound and observing that $\bar{\mathbf{X}}_i(\beta)$ is independent of $\bar{\mathbf{X}}_{i+2p}(\beta)$ for any $i \in [n]$, we split $\sum_i \bar{\mathbf{X}}_i(\beta)$ into $n/2p$ independent sums of r.v.s, we have

$$\mathbb{P}[\mathcal{E}_{\text{Net}}^c] \leq 2p |\mathcal{N}_\varepsilon| \cdot \mathbb{P} \left[\sum_{j=0}^{n/2p-1} (\bar{\mathbf{X}}_{2pj}(\beta) - \mathbb{E}[\bar{\mathbf{X}}(\beta)]) > \frac{9cn\theta^2}{p^2 \log n} \right],$$

where each summand has bounded variance and L^q -norm derived similarly as its expectation such that

$$\mathbb{E}\bar{\mathbf{X}}_i(\beta)^q \leq 2 \cdot \mathbb{P}[\langle s_i[\mathbf{x}], \beta \rangle \in [\lambda - B - M, \lambda + B + M]] \leq 2 \cdot \frac{1}{2\lambda} \cdot 2(B + M) \leq \frac{4c\theta^2}{p \log n},$$

and apply Bernstein inequality Lemma N.4 with $(\sigma^2, R) = (4c\theta^2 / (p \log n), 1)$, obtains

$$\begin{aligned} \mathbb{P} \left[\sum_{j=0}^{n/2p-1} (\bar{\mathbf{X}}_{2pj}(\beta) - \mathbb{E}[\bar{\mathbf{X}}(\beta)]) > \frac{9cn\theta^2}{p^2 \log n} \right] &\leq \exp \left[\frac{-(9cn\theta^2/p^2 \log n)^2}{2cn\theta^2/p^2 \log n + 2(9cn\theta^2/p^2 \log n)} \right] \\ &\leq \exp \left[\frac{-4cn\theta^2}{p^2 \log n} \right], \end{aligned}$$

thus when $n = Cp^5\theta^{-2} \log p$:

$$\mathbb{P}[\mathcal{E}_{\text{Net}}^c] \leq \exp \left[\log(2p) + 13p \log \left(\frac{2p \log n}{c \cdot c_\lambda} \right) - \frac{4cn\theta^2}{p^2 \log n} \right] \leq 1/n \quad (\text{L.37})$$

as long as $\frac{C}{\log C} > 10^5 / (c^2 \cdot c_\lambda)$. ■

M Algorithm for experiment

In our experiment, we use [Algorithm 1](#)(below), which is an adaptation of accelerated gradient descent [BT09] to the sphere. In particular, we apply momentum and increment by the Riemannian gradient via the exponential and logarithmic operators

$$\text{Exp}_{\mathbf{a}}(\mathbf{u}) := \cos(\|\mathbf{u}\|_2) \cdot \mathbf{a} + \sin(\|\mathbf{u}\|_2) \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|_2}, \quad (\text{M.1})$$

$$\text{Log}_{\mathbf{a}}(\mathbf{b}) := \arccos(\langle \mathbf{a}, \mathbf{b} \rangle) \cdot \frac{\mathbf{P}_{\mathbf{a}^\perp}(\mathbf{b} - \mathbf{a})}{\|\mathbf{P}_{\mathbf{a}^\perp}(\mathbf{b} - \mathbf{a})\|_2}, \quad (\text{M.2})$$

derived from [AMS09]. Here $\text{Exp}_{\mathbf{a}} : \mathbf{a}^\perp \rightarrow \mathbb{S}^{p-1}$ takes a tangent vector of \mathbf{a} and produces a new point on the sphere, whereas $\text{Log}_{\mathbf{a}} : \mathbb{S}^{p-1} \rightarrow \mathbf{a}^\perp$ takes a point $\mathbf{b} \in \mathbb{S}^{p-1}$ and returns the tangent vector which points from \mathbf{a} to \mathbf{b} .

Algorithm 1 SaS deconvolution with Accelerated Riemannian gradient descent

Require: Observation \mathbf{y} , sparsity penalty $\lambda = 0.5/\sqrt{p_0\theta}$, momentum parameter $\eta \in [0, 1)$.

Initialize $\mathbf{a}^{(0)} \leftarrow -\mathbf{P}_{\mathbb{S}^{p-1}} \nabla \varphi_\rho(\mathbf{P}_{\mathbb{S}^{p-1}}[\mathbf{0}^{p_0-1}; [\mathbf{y}_0, \dots, \mathbf{y}_{p_0-1}]; \mathbf{0}^{p_0-1}])$,

for $k = 1, 2, \dots, K$ **do**

 Get momentum: $\mathbf{w} \leftarrow \text{Exp}_{\mathbf{a}^{(k)}}(\eta \cdot \text{Log}_{\mathbf{a}^{(k-1)}}(\mathbf{a}^{(k)}))$.

 Get negative gradient direction: $\mathbf{g} \leftarrow -\text{grad}[\varphi_\rho](\mathbf{w})$.

 Armijo step $\mathbf{a}^{(k+1)} \leftarrow \text{Exp}_{\mathbf{w}}(t\mathbf{g})$, choosing $t \in (0, 1)$ s.t. $\varphi_\rho(\mathbf{a}^{(k+1)}) - \varphi_\rho(\mathbf{w}) < -t \|\mathbf{g}\|_2^2$.

end for

Ensure: Return $\mathbf{a}^{(K)}$.

N Tools

Lemma N.1 (Tail bound for Gaussian r.v.). *If $X \sim \mathcal{N}(0, \sigma^2)$, then its tail bound for $t > 0$ can be*

$$\mathbb{P}[X > t] \leq \frac{\sigma}{t\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (\text{N.1})$$

Lemma N.2 (Moments of the Gaussian random variables). *If $X \sim \mathcal{N}(0, \sigma^2)$, then ifor all integer $p \geq 1$,*

$$\mathbb{E}[|X|^p] \leq \sigma^p (p-1)!! \quad (\text{N.2})$$

Lemma N.3 (Gaussian concentration inequality). *Let $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ be a vector of n independent standard normal variables. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be an L -Lipschitz function. Then for all $t > 0$,*

$$\mathbb{P}[|f(\mathbf{x}) - \mathbb{E}f(\mathbf{x})| \geq t] \leq 2 \exp\left(-\frac{t^2}{2L^2}\right) \quad (\text{N.3})$$

Lemma N.4 (Moment control Bernstein inequality for scalar r.v.s). ([FR13], Theorem 7.30) *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent real-valued random variables. Suppose that there exist some positive number R and σ^2 such that $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{X}_i^2] \leq \sigma^2$ and*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[|\mathbf{x}_k|^p] \leq \frac{1}{2} \sigma^2 R^{p-2} p!, \quad \text{for all integers } p \geq 3.$$

Let $S \doteq \sum_{i=1}^n \mathbf{x}_i$, then for all $t > 0$, it holds that

$$\mathbb{P}[|S - \mathbb{E}[S]| \geq t] \leq 2 \exp\left(-\frac{t^2}{2n\sigma^2 + 2Rt}\right) \quad (\text{N.4})$$

Lemma N.5 (ε -net on sphere). [Ver10] Let (X, d) be a metric space and let $\varepsilon > 0$. A subset \mathcal{N}_ε of X is called an ε -net of X if for every point $x \in X$ there exists some point $y \in \mathcal{N}_\varepsilon$ so that $d(x, y) \leq \varepsilon$. There exists an ε -net \mathcal{N}_ε for the sphere \mathbb{S}^{n-1} of size $|\mathcal{N}_\varepsilon| \leq (3/\varepsilon)^n$.

Lemma N.6 (Hanson-Wright). [RV⁺13] Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent, subgaussian random variables with subgaussian norm $\sup_{p \geq 1} p^{-1/2} (\mathbb{E} |x_i^p|)^{1/p} \leq \sigma$. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$, then for every $t > 0$,

$$\mathbb{P} [|\mathbf{x}^* \mathbf{A} \mathbf{x} - \mathbb{E} \mathbf{x}^* \mathbf{A} \mathbf{x}| \geq t] \leq 2 \exp \left(-c \min \left(\frac{t^2}{64 \sigma^4 \|\mathbf{A}\|_F^2}, \frac{t}{8\sqrt{2} \sigma^2 \|\mathbf{A}\|_2} \right) \right). \quad (\text{N.5})$$

Lemma N.7 (Maximum of separable convex function). Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a convex function of the form $f(x) = x - s(x)$ with $s : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying

$$\frac{s(x)}{x} \leq \frac{s(y)}{y}, \text{ for all } x \geq y > 0.$$

Then for $n \in \mathbb{N}$ and $0 < N \leq nL$,

$$\max_{0 \leq \mathbf{x} \leq L, \|\mathbf{x}\|_1 \leq N} \sum_{i=1}^n f(\mathbf{x}_i) \leq N \left(1 - \frac{s(L)}{L} \right) \quad (\text{N.6})$$

Proof. Since the feasible set is a convex polytope; the convex function $\sum_{i=1}^n f(\mathbf{x}_i)$ is maximized at a vertex, and that its vertices consist of 0 and permutations of the vector $[\underbrace{L, \dots, L}_{\lfloor N/L \rfloor}, r, 0, \dots, 0]$, where $r =$

$N - \lfloor N/L \rfloor L \leq L$. Then the function value at the maximizing vector \mathbf{x}_* can be derived as:

$$\begin{aligned} \sum_{i=1}^n f(\mathbf{x}_{*i}) &= \lfloor \frac{N}{L} \rfloor f(L) + f(r) = \frac{N-r}{L} (L - s(L)) + (r - s(r)) \\ &= N \left(1 - \frac{s(L)}{L} \right) + r \left(\frac{s(L)}{L} - \frac{s(r)}{r} \right) \leq N \left(1 - \frac{s(L)}{L} \right) \end{aligned}$$

■

References

- [AMS09] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- [BC11] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 1st edition, 2011.
- [BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
- [CWB08] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.
- [FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.
- [GHJY15] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—Online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.
- [RV⁺13] Mark Rudelson, Roman Vershynin, et al. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18, 2013.
- [SQW17] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere ii: Recovery by riemannian trust-region method. *IEEE Transactions on Information Theory*, 63(2):885–914, 2017.
- [Ver10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.
- [WY13] Zaiwen Wen and Wotao Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1-2):397–434, 2013.